PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 16 November 15, 1930 Number 11

THE STRESS DISTRIBUTION IN WELDS

By William Hovgaard

Department of Naval Architecture, Massachusetts Institute of Technology

Communicated September 30, 1930

The distribution of stresses in welds, such as now applied more and more widely for the connection of structural members, is a problem which to the author's knowledge has not yet been solved. The nearest approach to a solution was made by Professor Cyril Batho who proposed formulas for calculating the partition of the load in riveted joints dealing separately with each rivet.[1] The solution here offered depends on a process of integration and variation, which presupposes continuity, such as often exists in welds, but it applies also with fair approximation where the weld is intermittent and may be used in some cases of riveted connection. It is valid only within the limit of elasticity.

Space does not permit a discussion of any of the numerous special cases, such as overlapped or butt-strapped joints, stiffeners attached to bulkheads or other surfaces subject to pressures, etc. In this paper will be discussed only the simple case where a bar is connected, generally as a reenforcement, to another structural member which is subject to simple tension or compression. The latter member is supposed to be of greater length than the former. This case is fundamental, since the formulas that apply to such a simple structure are applicable with certain modifications to all other cases of welds.

In order to fix ideas, consider a plate which forms an integral part of a larger structure subject to elongation. A bar of any form or section, extending in the direction of the strain over a certain length of the plate, is connected to it by a continuous weld and is thus, in virtue of its connection with the plate, forced to elongate. The tensile stresses in the plate are communicated to the bar through shearing stresses in the weld. So long as the plate is unstrained, the bar is inert, but when the plate elongates the bar comes to act as a reënforcement. In practice we find an approach to this condition in many cases, as for instance in cover plates on the chords or flanges of girders, in the intermittent longitudinal girders fitted in the bottom and under the decks of certain ships and in all local reën-

forcements. We refer to the two members here under consideration as the plate and the bar, but each of these members may take many different forms without impairing essentially the validity of the following treatment. In the diagram figure 1, AB represents the plate and CD the bar. We take O at the middle of the bar as origin and place the axis OX along OB. In order to reduce the problem to its simplest form, we imagine the bar to be a simple doubling plate as in figure 2a, or it may be a plate strip placed normal to a plate and welded to it as indicated in section in figure 2b. We assume that bending due to unsymmetrical loading is prevented by adjacent structures, such as is usually the case in practice.

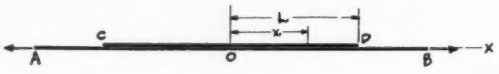


FIG. 1

FIG. 2

It is at once clear that by symmetry there can be no shearing stress in the weld at O and we need only consider one side of the structure, say, from O to D. When the plate elongates, the bar will resist the motion and each point of the bar, with the exception of O, will be a little displaced relative to the corresponding point in the plate, the displacement increasing from zero at O to a maximum at the end of the bar. The shearing stresses will likewise increase from the middle of the bar toward the end and we now make the assumption that *the displacement at every point is proportional to the shearing stress in the weld at that point.*

The distribution of the stresses in a given bar depends on the nature of the welding material and on the shearing area. When the weld metal yields readily, the elongation and hence the stress in the bar at the middle will be considerably smaller than in the plate, but when the weld metal is more stiff and unyielding, the stress in the bar at O will increase and approach that in the plate at that point. In the limit, when the weld is very rigid, the bar will be under the same stress as the plate except at the ends where the shearing stress will rise steeply to a high maximum.

The problem is to determine the law according to which the stresses vary in the plate and the bar as well as in the weld.

The conditions of equilibrium furnish two equations between the tensile stresses p_x in the plate, $\bar{p}_x$ in the bar and the shearing stress q_x in the weld. A third equation, referred to as the "displacement equation," is obtained

on the basis of the assumed relations between shearing stress and displacement stated above. The condition of symmetry, that $q = 0$ at the middle of the bar, furnishes a fourth equation, and a fifth equation is found on the principle that the total elastic work in the structure between sections through O and D shall be a minimum. The integral expressing this work contains the stress q which is an unknown function of x, the form of which is determined by the method of variation. This method leads to a differential equation of the second order, the solution of which contains two unknown constants.

We have thus five equations for the determination of the three stresses and the two constants of integration.

In the Appendix is given the fundamental mathematical solution, but space does not permit a discussion of its application to practical problems.

Figure 3 shows curves of stress calculated from the formulas, assuming various values of the "displacement coefficient," μ, which connects the shearing stress with the displacement. Experiments are now in progress at the Massachusetts Institute of Technology in order to determine the value of this coefficient.

APPENDIX

Fundamental Solution.—Consider a bar welded to a plate as described above and as indicated in figures 1 and 2. The plate is subject to a tensile stress p beyond the ends of the bar but in way of the bar the stress is reduced. We use the following notation:

$2L$ = total length of bar
A = sectional area of plate
a = sectional area of bar
α = sectional area of weld per unit length of bar, assumed to be constant
q_x = shearing stress in weld at the point x
p_x = average tensile stress in plate across a section at x
$\bar{p}_x$ = average tensile stress in bar across a section at x
μq_x = average linear displacement of a section of the bar relative to the corresponding section of the plate at x.

Actually the stresses are very unevenly distributed over a transverse section in the plate as well as in the bar, but the average elongation or contraction is taken approximately equal to that which would be caused by the average stress. The displacement is due partly to shearing in the weld material, partly to strains in the bar and in the plate. The action is very complex but the average displacement is here assumed to be proportional to the shearing stress so that μ is a constant, which we call the "displacement coefficient."

The average stress across any section of the plate must be equal to the original stress p minus the average stress due to the pull of the weld outside that section:

$$p_x = p - \frac{\alpha}{A}\int_x^L q dx. \tag{1}$$

The average stress across any section of the bar is entirely due to the pull of the weld outside that section:

$$\bar{p}_x = \frac{\alpha}{a}\int_x^L q dx. \tag{2}$$

The average displacement of the bar at x must be equal to the difference between the elongation of the plate and the bar from O to x:

$$\mu E q_x = px - \frac{\alpha(A+a)}{aA}\int_0^x \left[\int_x^L q dx\right] dx. \tag{3}$$

The total elastic work, W, consists of three parts, one due to the shearing stress forces, one due to the tension in the plate and one due to the tension in the bar. Hence we obtain

$$W = \int_0^L \left\{\frac{\alpha\mu}{2} q_x^2 + \frac{A}{2E}\left[p - \frac{\alpha}{A}\int_x^L q dx\right]^2 + \frac{\alpha^2}{2aE}\left[\int_x^L q dx\right]^2\right\} dx. \tag{4}$$

It is required to determine q as a function of x so as to make this integral a minimum, but since q occurs not only explicit but also involved in an integral, the latter is best dealt with by regarding it as another dependent variable. We write:

$$\int_x^L q dx = z. \tag{5}$$

Hence z is now an unknown function of x connected with q by equation (5), which by differentiation may be given the form of a differential equation:

$$\frac{dz}{dx} = -q_x. \tag{6}$$

This may be incorporated as an equation of condition in the integral for W after multiplication with an indeterminate quantity λ, itself a function of x. Thus we obtain the integral:

$$W = \int_0^L \left\{\frac{\alpha\mu}{2} q^2 + \frac{A}{2E}\left(p - \frac{\alpha z}{A}\right)^2 + \frac{\alpha^2 z^2}{2aE} + \lambda\left(q + \frac{dz}{dx}\right)\right\}. \tag{7}$$

The variation of W, when both q and z are varied and after reduction of

$\frac{d\delta z}{dx}$ gives the following expression which must be equal to zero if W is to be a minimum.

$$\delta W = \int_0^L \left\{ [\alpha\mu q + \lambda]\delta q - \left[\frac{\alpha}{E}\left(p - \frac{\alpha z}{A}\right) - \frac{\alpha^2 z}{aE} + \frac{d\lambda}{dx}\right]\delta z \right\} dx = 0. \quad (8)$$

Since δq and δz are now entirely independent and arbitrary, this equation can be true only if each of the factors of δq and δz vanish. Hence,

$$q_x = -\frac{\lambda}{\alpha\mu} \text{ and therefore from (6): } \frac{dz}{dx} = \frac{\lambda}{\alpha\mu} \quad (9)$$

and

$$\frac{d\lambda}{dx} = \frac{\alpha^2(A+a)z}{aAE} - \frac{\alpha p}{E}. \quad (10)$$

Differentiating (10) we get:

$$\frac{d^2\lambda}{dx^2} = \frac{\alpha^2(A+a)}{aAE}\frac{dz}{dx} = \frac{\alpha(A+a)\lambda}{aAE\mu},$$

or:

$$\frac{d\lambda^2}{dx^2} - \frac{\alpha(A+a)}{aAE\mu}\lambda = 0. \quad (11)$$

Writing

$$\sqrt{\frac{\alpha(A+a}{aAE\mu}} = m, \quad (12)$$

the solution of (11) is:

$$\lambda = Re^{mx} + Se^{-mx}, \quad (13)$$

and therefore:

$$q_x = -\frac{1}{\alpha\mu}(Re^{mx} + Se^{-mx}), \quad (14)$$

This is the general solution of the problem, and it remains only to determine the constants of integration. The coefficient m is known, provided we are able to estimate the value of the displacement coefficient μ.

At the middle of the bar there is no shearing stress and hence when $x = 0$, $q = 0$. At the end of the bar, where $x = L$ we have the greatest shearing stress, which we denote by q_L.

Substituting in (14) and introducing hyperbolic functions, we find:

$$R = -\frac{\alpha\mu q_L}{2\sinh mL} \text{ and } S = -R,$$

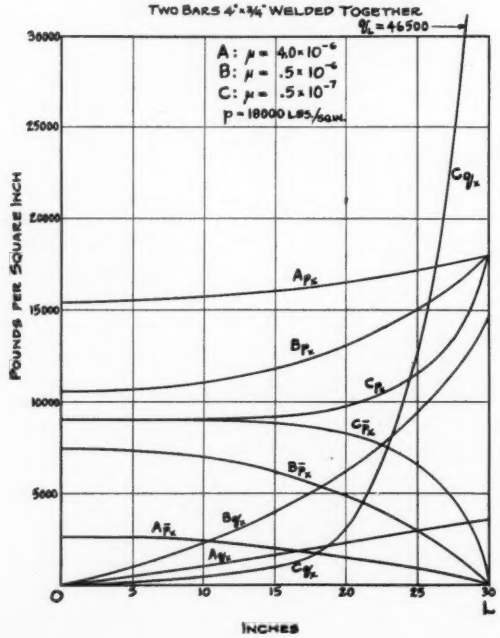
which on substitution in (14) gives:

$$q_x = \frac{q_L \sinh mx}{\sinh mL}. \tag{15}$$

In equation (3) substitute this value of q_x and put $x = L$ in the upper limit of the principal integral. After integration solve for q_L:

$$q_L = \frac{p \tanh mL}{m\mu E}, \tag{16}$$

FIG. 3
CURVES FOR CALCULATED STRESSES
TWO BARS 4" × ¾" WELDED TOGETHER



and therefore the shearing stress at any point is given by the following simple expression:

$$q_x = \frac{p}{m\mu E} \frac{\sinh mx}{\cosh mL}. \tag{17}$$

The stress in the bar at any point is found from (2):

$$\bar{p}_x = \frac{pA}{A + a} \frac{\cosh mL - \cosh mx}{\cosh mL} \tag{18}$$

from which is obtained the stress at the middle:

$$\bar{p}_0 = \frac{p\,A}{A + a} \frac{\cosh mL - 1}{\cosh mL}. \tag{19}$$

The stress in the plate at any point is then found from

$$p_x = p - \frac{a}{A} \bar{p}_x. \tag{20}$$

Figure 3 gives curves of stress for a steel bar five feet long and 4 in. by $^3/_4$ in. in section welded to another steel bar of the same section but of greater length. The latter bar, representing the "plate" as defined above, is subjected to a tensional stress of 18,000 lbs. per sq. in. There are three sets of curves marked A, B and C corresponding to various values of the displacement coefficient, each set comprising one curve for q_x, one for $\bar{p}_x$ and one for p_x. The rapid rise of the curve for the shearing stress $\bar{q}_x$ at the end of the bar, when the displacement coefficient is small, is of interest. It indicates a concentration of shearing stress at the end when the welded connection is very stiff and unyielding. In that case the tensile stress in the bar is equal to that in the plate up to nearly midway between O and L.

[1] *J. Frank. Inst.*, **182**, No. 5, November, 1916.

THE STRESS DISTRIBUTION IN WELDED OVERLAPPED JOINTS

BY WILLIAM HOVGAARD

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated October 16, 1930

In a previous paper[1] the writer described a general method of determining stresses in welds. The present paper deals with the special case where overlapping plates are connected by welds running parallel to the line of pull.

Consider a double-strapped joint as shown in figure 1, where the main plates are of equal thickness. Since there is symmetry about a transverse plane through O, we need only consider one side of the joint, such as from O to D. The straps are supposed to be connected by fillet welds along the sides, but not at the transverse end lines, so that there are altogether four lines of weld, each of length $2L$. In the theoretical treatment of this joint it is imagined that the two straps are replaced by one of twice the thickness but the total effective area of the welds is supposed to be the same as in the double strap. We have then on one side of the butt line the equivalent of a simple overlapped joint as shown in figure 2 where plate (2) takes the place of the straps and plate (1) represents one of the plates. It is not

asserted that the resulting formulas actually represent conditions in single overlaps, since, on account of unsymmetry, bending stresses necessarily occur in such joints, but it is believed that when an overlapped joint is

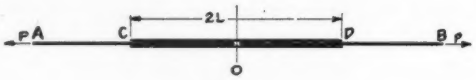


FIG. 1

prevented from bending by adjacent structures, as frequently happens, the formulas will give a good approximation to the stresses also in that case. Figure 3 shows the overlap in plan view and section, indicating the two lines of weld.

FIG. 2

Mathematical Solution.—The notation is, with certain modifications, the same as in the previous paper.

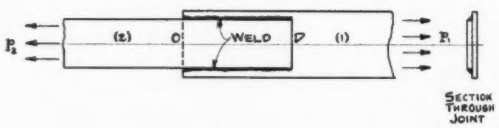


FIG. 3

L is here the length of the simple overlap represented in figures 2 and 3.

A = sectional area of plate (1).
a = sectional area of plate (2) which replaces the straps.
α = the sum of the sectional areas of the welds per unit length of the overlap, hence equal to twice the area per unit length of each of the welds in figure 3.
p_{x1} = average tensile stress in plate (1) across a section at x.
p_{x2} = average tensile stress in plate (2) across a section at x.

Again

q_x = shearing stress in weld at the point x.
μq_x = average linear displacement of a section of plate (1) relative to the corresponding section of plate (2). The coefficient μ is assumed to be constant in any given joint, but may vary in different joints.

The plates are subject to a pull P uniformly distributed over the sectional area, so that we have the stresses in the plates outside the overlap:

$$p_1 = \frac{P}{A} \text{ and } p_2 = \frac{P}{a}. \tag{1}$$

The average stress in plate (1) at any point x in the overlap is equal to p_1 minus the average stress due to the pull of the weld between x and L.

$$p_{x1} = p_1 - \frac{\alpha}{A}\int_x^L q\,dx. \tag{2}$$

The average stress in plate (2) at the same point is due entirely to the pull of the weld outside x:

$$p_{x2} = \frac{\alpha}{a}\int_x^L q\,dx. \tag{3}$$

We have obviously:

$$\alpha\int_0^L q\,dx = P = p_1 A = p_2 a. \tag{4}$$

The total elastic work in the joint from O to D is expressed as in the previous paper:

$$W = \int_0^L \left\{\frac{\alpha\mu q^2}{2} + \frac{A}{2E}\left(p_1 - \frac{\alpha}{A}\int_x^L q\,dx\right)^2 + \frac{\alpha^2}{2aE}\left[\int_x^L q\,dx\right]^2\right\} dx. \tag{5}$$

It is required to determine q so as to make W a minimum. Making again $\int_x^L q\,dx = z$ and proceeding as in the previous paper, we arrive at the same equation:

$$q_x = -\frac{1}{\alpha\mu}(Re^{mx} + Se^{-mx}). \tag{6}$$

Here R and S are constants of integration and

$$m = \sqrt{\frac{\alpha(A+a)}{aA\mu E}}. \tag{7}$$

When $x = 0$, $q_x = q_0$ and when $x = L$, $q_x = q_L$ from which after substitution in (6) we obtain:

$$R = \frac{\alpha\mu}{2\sinh mL}(q_0 e^{-mL} - q_L) \text{ and } S = -\frac{\alpha\mu}{2\sinh mL}(q_0 e^{mL} - q_L)$$

$$\therefore q_x = \frac{q_0 \sinh m(L-x) + q_L \sinh mx}{\sinh mL}. \tag{8}$$

Substitute this value of q in (3), carry out the integration and find the stress in plate 2:

$$p_{x2} = \frac{\alpha}{am \sinh mL} [q_L(\cosh mL - \cosh mx) + q_0(\cosh m(L - x) - 1)]. \tag{9}$$

When $x = 0$, the stress in plate 2 at O is:

$$p_{02} = \frac{\alpha(q_0 + q_L)(\cosh mL - 1)}{am \sinh mL}. \tag{10}$$

The displacement of plate 2 relative to plate 1 at the right-hand end of the joint D is found as the difference in elongation of the two plates from O to D plus the displacement at O:

$$E\mu q_L = \int_0^L p_{x1}\,dx - \int_0^L p_{x2}\,dx + \mu q_0 E. \tag{11}$$

Substituting from (2), (3), (8) and observing that according to (7):

$$\frac{\alpha(A + a)}{aA} = m^2\mu E,$$

we find:

$$q_L = \frac{p_1 \tanh mL}{m\mu E} + \frac{q_0}{\cosh mL}. \tag{12}$$

Since $p_{02} = p_2$ we have from (10)

$$q_0 + q_L = \frac{amp_2 \sinh mL}{\alpha(\cosh mL - 1)}, \tag{13}$$

and then from (12) and (13):

$$q_L = \frac{p_1 \sinh mL}{m\mu E(1 + \cosh mL)} + \frac{amp_2}{\alpha \sinh mL}, \tag{14}$$

$$q_0 = -\frac{p_1 \sinh mL}{m\mu E(1 + \cosh mL)} + \frac{amp_2 \cosh mL}{\alpha \sinh mL}. \tag{15}$$

When the plates are of equal sectional area, $a = A$, we have $q_L = q_0$.

The stress in plate 2 is now found from (9) and in plate 1 from:

$$p_{x1} = p_1 - \frac{a}{A} p_{x2}. \tag{16}$$

Test Results.—The formulas given above have unexpectedly found verification in a series of unique and valuable tests carried out at the University of Pittsburgh by Professor J. Hammond Smith[2] and published in the *Journal of the American Welding Society.*

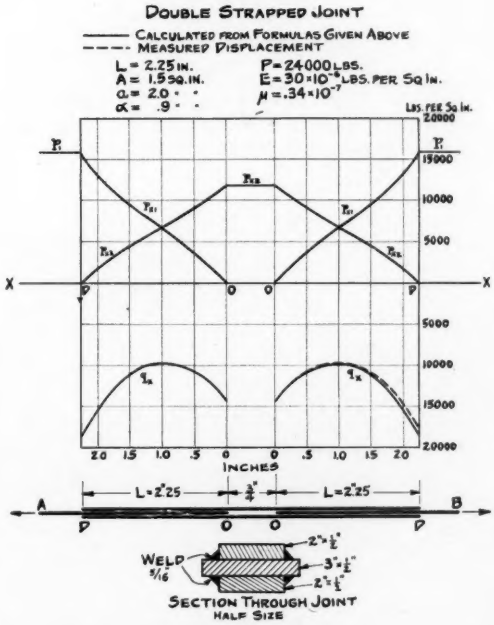
These tests were made on double strapped plates, the dimensions of

which in most cases were $3'' \times 1/2''$, while the straps were $2''$ wide and of thickness varying from $3/8''$ to $3/4''$. A few tests were made on plates and straps of greater width. Each strap was connected by two lengthwise fillet welds to the respective plate.

The displacement of the straps relative to the plates were measured with great accuracy at a number of sections and were plotted in curves, with the axis of abscissae parallel with the welds.

No analysis of the tests was given in the report, but the writer found a remarkably close correspondence between the curves, which represent

FIG. 4
CURVES OF STRESS
DOUBLE STRAPPED JOINT



the observed displacement μq and the curves for q calculated from formula (8), as seen from figure 4. For all the narrow specimens, it was found that after proper scale adjustment these curves were practically identical, showing that μ was actually constant for each specimen. Some variation in μ was found between different specimens.

In case of the wider specimens, the correspondence was not so good, presumably because here the measurements, which were taken close to the lines of weld only, did not so well represent the average displacement of the whole cross section as assumed in the theory.

Figure 4 shows the curves for shearing stress and for tensile stress calculated according to the formulas for one of the test specimens, marked 1B in Professor Hammond Smith's report. The dimensions of the plates and the straps as well as other data pertaining to the test are given in the figure. For the sake of clearness the curves for the tensile stresses are laid off above the axis of abscissae and the curves for shearing stress below the axis. There is added a curve (dotted) for the measured displacements as obtained from figure 10 of the report, reduced to such a scale that the ordinate at O is the same as in the curve for the calculated value of q. The close correspondence between these two curves, one representing q, the other μq, is obvious, and is found also for the other test pieces of narrow width.

The value of the displacement coefficient μ in those tests was found to vary from 0.3×10^{-7} to 0.4×10^{-7}. This coefficient is likely to be different in joints of different kind and as yet little is known of its value in various cases, but it is hoped that experiments now in progress at the Massachusetts Institute of Technology will throw further light on this subject.

[1] These PROCEEDINGS, November, 1930, p. 667.

[2] Report submitted to the Fundamental Research Committee of the American Bureau of Welding, *J. Am. Weld. Soc.*, **8**, Sept., 1929; "Stress Strain Characteristics of Welded Joints."

A RADIO SYSTEM FOR BLIND LANDING OF AIRCRAFT IN FOG

BY H. DIAMOND AND F. W. DUNMORE

NATIONAL BUREAU OF STANDARDS, WASHINGTON, D. C.

Read before the Academy, September 19, 1930

This sytem includes three elements in order to indicate the position of the aircraft in three dimensions as it approaches and reaches the point of landing.

1. *Lateral Position.*—Such position given for the purpose of keeping the airplane directed to and over the runway is indicated by two vibrating reeds on the pilot's instrument board, the driving electromagnets of which are connected to the output of the airplane's radio receiver. These reeds, one of which is mechanically tuned to 65 cycles and the other to 86.7 cycles, are actuated by a radio signal sent from two coil antennas crossed at 90°, the signal from one coil antenna being modulated at 65 cycles and that from the other at 86.7 cycles. On the course (i.e., along the line bisecting the angle between the two antennas) the reed vibration amplitudes are equal. Off the course they are unequal, the reed vibrating with the greater

amplitude being on the side to which the airplane has deviated. An automatic volume control feature is used to keep the reed amplitudes within bounds as the field is approached.

2. *Vertical Guidance.*—A high-frequency (100 megacycles, 3 meters)

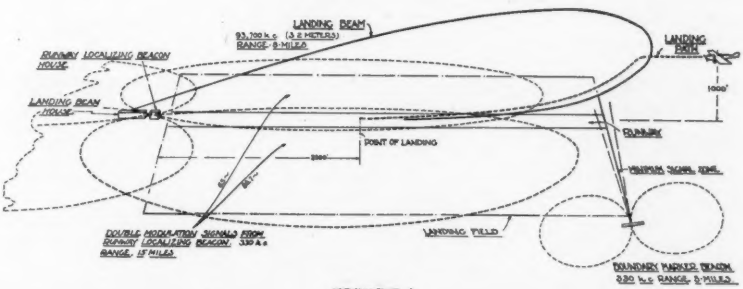


FIGURE 1

Three dimensional view showing radio system of blind landing aids.

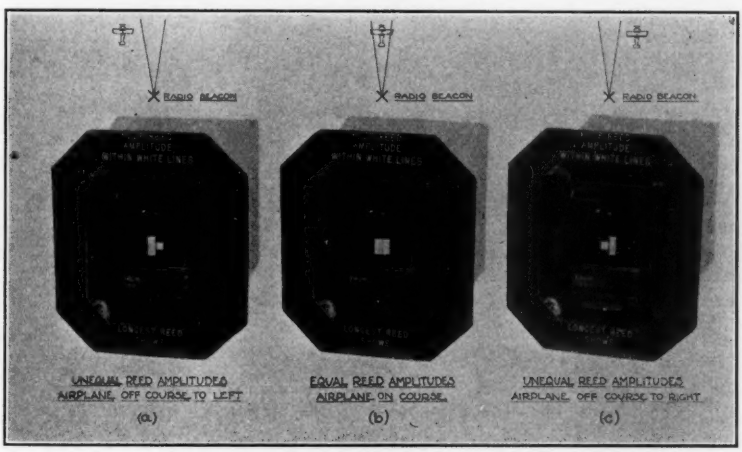


FIGURE 2

The tuned reed indicator used for giving horizontal guidance.

beam directed over the runway at an angle of 8° above the horizontal and located at the further end of the landing field is used for such guidance.

On the airplane, the signal current in the output circuit of the special high-frequency receiving set employed is rectified and passed through a d. c. microammeter mounted on the instrument board. The airplane does

not fly on the axis of the beam, but on a curved path under the beam whose curvature diminishes as the ground is approached. The path is the line of equal intensity of received signal below the axis of the beam. The diminution of intensity as the airplane drops below the beam axis is compensated by the increase of intensity due to approaching the beam transmitter. Thus, by flying the airplane along such a path as to keep the de-

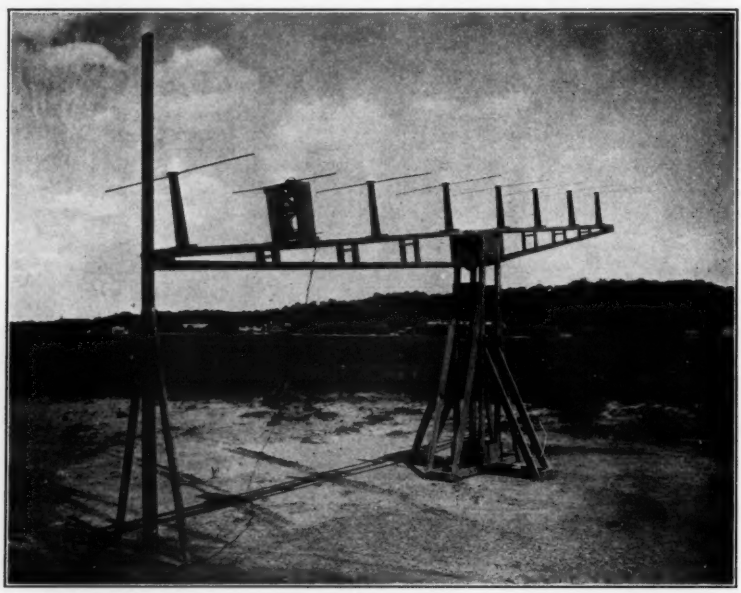


FIGURE 3

Antenna array for directing the ultra-high frequency landing beam used for vertical guidance.

flection of the microammeter on the instrument board constant, the pilot comes down to ground on a curved line suitable for landing.

No manipulations on the part of the pilot are required. The tuning is fixed. Since a line of constant field intensity is followed no control of volume is necessary.

3. *Longitudinal Guidance.*—A field boundary marker beacon is used for this purpose. It operates on the same carrier frequency as the runway localizing beacon and both beacons are received simultaneously on the medium-frequency receiving set. A modulation frequency of 1000 cycles is used. A coil antenna oriented to give a minimum signal zone along the

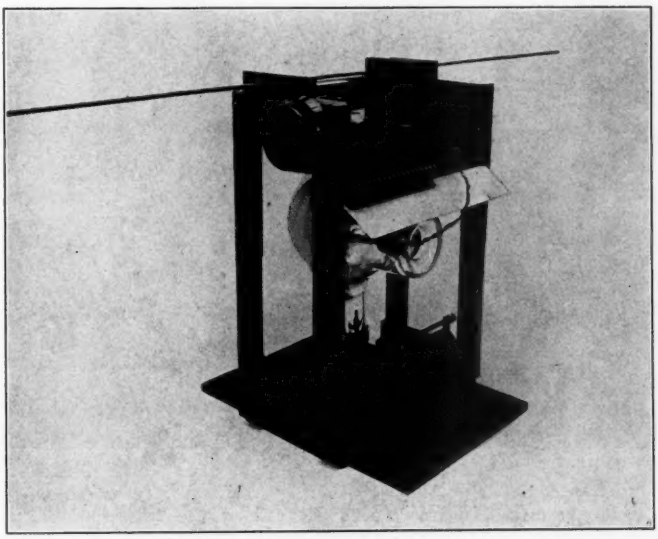


FIGURE 4

The ultra-high frequency transmitting tube and associated circuit.

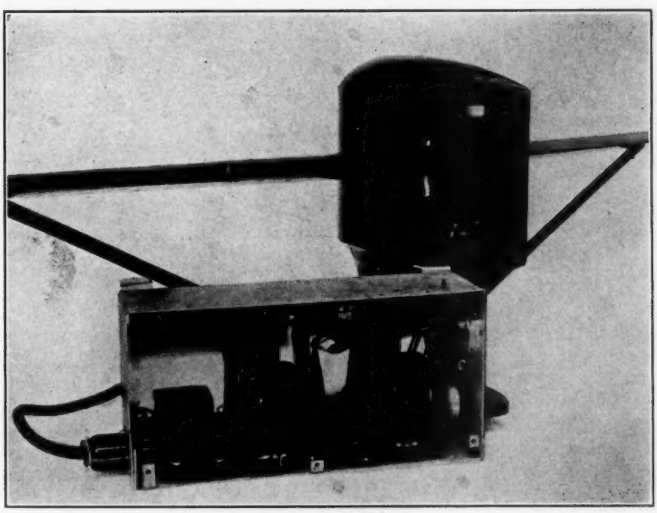


FIGURE 5

Equipment for receiving the ultra-high frequency landing beam signals.

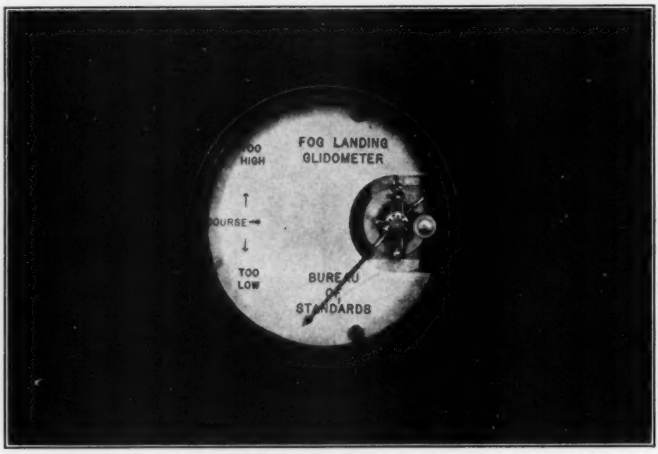


FIGURE 6

Landing beam indicator used on the pilot's instrument board to show the relative position of the airplane with respect to the proper landing path.

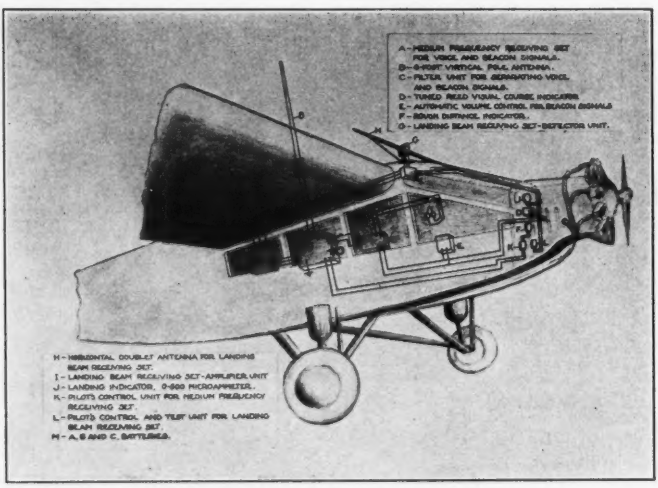

A - MEDIUM FREQUENCY RECEIVING SET FOR VOICE AND BEACON SIGNALS.
B - 6-POST VERTICAL POLE ANTENNA.
C - FILTER UNIT FOR SEPARATING VOICE AND BEACON SIGNALS.
D - TUNED REED VISUAL COURSE INDICATOR.
E - AUTOMATIC VOLUME CONTROL FOR BEACON SIGNALS
F - SONIC DISTANCE INDICATOR.
G - LANDING BEAM RECEIVING SET-DETECTOR UNIT.
H - HORIZONTAL DOUBLET ANTENNA FOR LANDING BEAM RECEIVING SET.
I - LANDING BEAM RECEIVING SET-AMPLIFIER UNIT
J - LANDING INDICATOR, 0-500 MICROAMMETER.
K - PILOT'S CONTROL UNIT FOR MEDIUM FREQUENCY RECEIVING SET.
L - PILOT'S CONTROL AND TEST UNIT FOR LANDING BEAM RECEIVING SET.
M - A, B AND C, BATTERIES.

FIGURE 8

Skeleton view of Bureau of Standards airplane showing complete receiving apparatus for blind landing system.

border of the field is used, the minimum signal indicating to the pilot that he is over the edge of the field.

Figure 1 outlines the complete landing system. The airplane is kept horizontally in line with, and over the runway, by means of the 65 and 86.7 cycle signals shown by the light-dotted ellipses. Along the line of the runway these two signals are of equal strengths giving equal reed amplitudes. The heavy solid line indicates the high frequency beam oriented vertically at 8° to the horizontal. The heavy dotted line shows the vertical landing path as followed by the airplane, when the pilot keeps the received signal at a constant value. As the source is approached it is necessary to continually drop under and out of the beam to hold a constant received signal.

Figure 2 shows the tuned reed indicator for giving horizontal guidance. The three views show how the reed indicator looks to the pilot when on course and off course to the right or left.

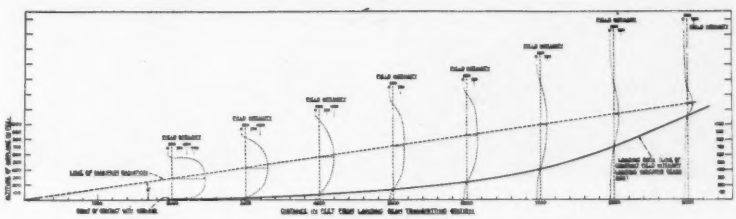


FIGURE 7

True landing path as determined experimentally.

Figure 3 shows the antenna array for directing the 100 megacycle (3 meters) beam at an angle of 8° with the horizontal. A 500-watt transmitting tube is used.

Figure 4 is a close-up view of the 500-watt tube and associated circuit. To adjust this circuit to the desired frequency a dial is provided which moves one of the plates of the air condenser shown above the tube.

Figure 5 is a view of the equipment for receiving the 100-megacycle signal. The detector tube may be seen in the stream-lined box. The horizontal doublet antenna is in the horizontal stream-lined wooden section. The one-stage audio amplifier and rectifier unit is shown below. This unit is placed inside the airplane.

Figure 6 shows the indicating instrument which is connected to the output of the amplifier-rectifier unit just shown. This is a 0–500 d.c. microammeter, and the needle is held in the horizontal position when the airplane is following the landing path. If the airplane goes above the landing path the signal increases and the needle rises. If it goes below, the reverse is true.

FIGURE 9

Instrument board on Bureau of Standards' airplane.

FIGURE 10

Vertical pole antenna and horizontal doublet antenna used in reception of blind landing system signals.

Figure 7 shows the true landing path as determined experimentally. Here the landing glide starts at 1100 feet altitude, 9000 feet from the landing beam transmitting set, the airplane coming to rest on the ground 2000 feet in front of the beam transmitting set. It will be noted that the landing indicator reads 250 ($^1/_2$ scale) during the whole of the glide.

Figure 8 is an inside view of the Bureau of Standards airplane showing the location of the blind landing apparatus.

Figure 9 shows the instrument board of the airplane. The tuned reed indicator is shown at the top and the landing beam indicator at the right center.

Figure 10 is a view of the airplane showing the vertical pole antenna for receiving the runway localizer signal for horizontal guidance and the horizontal doublet antenna for receiving the high-frequency beam signal for vertical guidance.

A SENSITIVE INDUCTION BALANCE FOR THE PURPOSE OF DETECTING UNEXPLODED BOMBS

BY THEODORE THEODORSEN

THE NATIONAL ADVISORY COMMITTEE FOR AERONAUTICS

Read before the Academy April 29, 1930

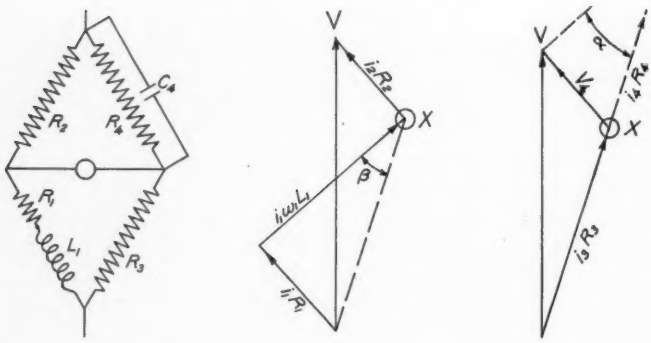
Introduction.—In connection with the construction of a channel for the testing of seaplane floats at the Langley Field laboratories of the National Advisory Committee for Aeronautics, the Committee was recently confronted with the problem of locating unexploded bombs buried in the earth.

The site of the channel was located in close proximity to an area which had previously been used by the Army Air Corps for target practice.

The Committee made a careful survey of the situation and requested information from all organizations which it was believed might be able to render assistance. None of the proposed methods proved to be satisfactory and had to be abandoned.

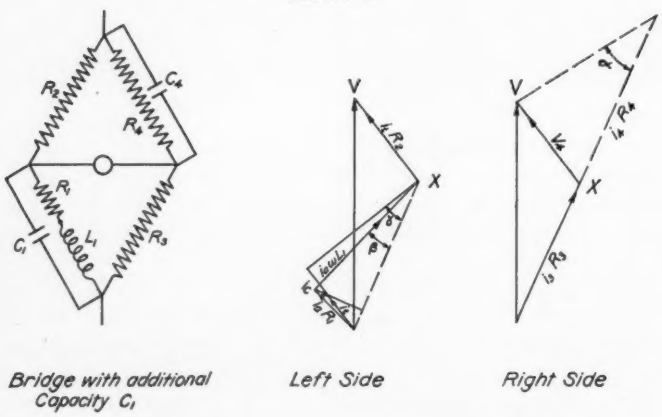
The Laboratory, after trying various proposed schemes, succeeded in devising an instrument which possessed the necessary properties. This instrument is of a very simple design and is very convenient in operation.

The entire site of the channel has been scanned with this detector with good results. The instrument is perfectly fool-proof and requires, for its operation, no skilled operators. It was felt that this paper might be of interest to those confronted with similar problems. The question is also of some theoretical interest, and it is this aspect of the problem which will be taken up in the present paper.



VECTOR DIAGRAM OF MAXWELL BRIDGE IN BALANCE

FIGURE 1



THE EFFECT OF CAPACITY IN THE INDUCTANCE ARM OF A MAXWELL BRIDGE

FIGURE 2

Analysis of Possible Methods of Detection of Metallic Bodies Buried in the Earth.—The line of study followed during the development of the N. A. C. A. detector will be indicated in the following:

It is probable that the most conspicuous property of a metallic body, as distinguished from that of the soil, is its effect on a magnetic field.

If a coil of a diameter which is comparable in magnitude to the distance desired is brought close to the surface above the buried metal, a measurable change in the magnitude of its self-inductance takes place. The power factor or phase angle is also, in general, changed, due to the eddy currents set up within the metal. Certain relations should exist between the dimensions of the coil and those of the object for greatest sensitivity.

A rough estimate will show that the change to be expected is so minute that the most refined methods of detection must be resorted to. One possibility would be to employ a beat-frequency oscillator. The search coil could form the inductance of one of the oscillating circuits. It was conceived, however, that it would be difficult to insure sufficient stability.

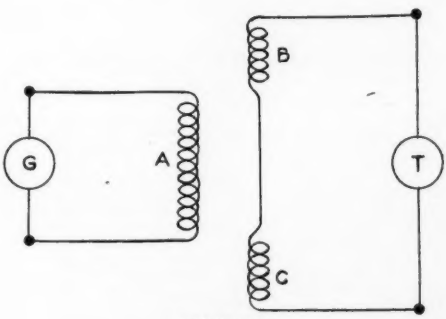


FIGURE 3

Methods for detecting changes in the self-inductance by means of bridges were considered next.

The Bureau of Standards has given some attention to this question during the war. A brief résumé of such work is published in the "War Works" of the Bureau.[1]

The Bureau recommended the use of a Maxwell bridge in connection with a large search coil of several Henrys, as being possibly the most sensitive method. The Committee tried this scheme.

It was found that the arrangement was entirely too complicated for use outside of the Laboratory. It was also found that any great sensitivity could not be obtained for theoretical reasons. The Maxwell bridge is shown in figure 1. The conditions for balance may be given most easily by means of vector diagrams, as shown in the same figure.

We obtain the following relations:

$$V_1 = I_1 R_1 \frac{1}{\sin \beta} \quad V_2 = I_1 R_2 \quad tg\alpha = \frac{1}{R_4 \omega C_4}$$

$$V_3 = I_3 R_3 \quad V_4 = I_3 R_4 \sin \alpha \; tg\beta = \frac{R_1}{\omega L_1}.$$

Conditions for zero bridge current are:

$$V_1 = V_3, \; V_2 = V_4 \text{ and } \alpha = \beta$$

or:

$$R_1 R_4 = R_2 R_3 = \frac{L_1}{C_4}.$$

It appears that the balancing of the bridge is independent of the frequency. This, however, is only correct for a true Maxwell bridge.

As a matter of fact, the actual bridge is remote from this ideal condition. The resistances contain both capacity and inductance. The condenser is not free from leakage, and, in particular, the inductance L_1 contains considerable self-capacity.

Figure 2 shows the actual bridge with this latter effect taken into account. The capacity of the coil L_1 may very nearly be considered as shunted across the coil as far as its effect is concerned.

We obtain the following relations as conditions of balance:

$$V_1 = V_3,\ V_2 = V_4 \text{ and } \alpha = \gamma$$

where:

$$V_1 = I_a R_1 \frac{1}{\sin \beta} \quad V_3 = I_3 R_3 \quad V_2 = I_i R_2 \quad V_4 = I_3 R_4 \sin \alpha$$

$$tg\alpha = \frac{1}{R_4 \omega C_4} \quad tg\gamma = \frac{\frac{V_1}{R_1} \sin^2 \beta}{\frac{V_1}{R_1} \cdot \cos \beta \sin \beta - V_1 \omega C_1} = \frac{1}{\frac{1}{tg\beta} - \omega C_1 R_1 \left(1 + \frac{1}{tg^2 \beta}\right)}.$$

but $tg\beta = \dfrac{R_1}{\omega L_1}$ which gives: $tg\gamma = \dfrac{1}{\dfrac{\omega L_1}{R_1} - \omega R_1 C_1 \left(1 + \left(\dfrac{\omega L_1}{R_1}\right)^2\right)}$.

The condition $\alpha = \gamma$ gives:

$$R_4 \omega C_4 = \frac{\omega L_1}{R_1} - \omega R_1 C_1 \left(1 + \left(\frac{\omega L_1}{R_1}\right)^2\right)$$

or

$$R_1 R_4 = \frac{L_1}{C_4} - \frac{R_1^2 C_1}{C_4} \left(1 + \left(\frac{\omega L_1}{R_1}\right)^2\right). \qquad \text{I}$$

The first condition gives:

$$I_a R_1 \frac{1}{\sin \beta} = I_3 R_3 \text{ but } I_a \sin \beta = I_i \sin \gamma \text{ and } I_i \frac{\sin \gamma}{\sin^2 \beta} R_1 = I_3 R_3.$$

The second condition gives:

$$I_i R_2 = I_3 R_4 \sin \alpha$$

Consequently:

$$\frac{I_i}{I_3} = \frac{R_3}{R_1} \frac{\sin^2 \beta}{\sin \gamma} = \frac{R_4 \sin \alpha}{R_2} \text{ but } \gamma = \alpha \text{ from } I, \text{ and}$$

$$R_1R_4 = R_2R_3\frac{\sin^2\beta}{\sin^2\alpha} = R_2R_3 \cdot \frac{1 + R_4^2\omega^2C_4^2}{1 + \left(\frac{\omega L_1}{R_1}\right)^2}. \qquad \text{II}$$

Equation I becomes independent of ω only if: $C_1 = 0$. In this case: $R_1R_4 = \frac{L_1}{C_4}$ which is the Maxwell relation of balance.

Equation II is independent of ω if: $R_4C_4 = \frac{L_1}{R_1}$. Then, also, $R_1R_4 = R_2R_3$ which is the other Maxwell equation.

The relations I and II may be written:

$$f(\omega R_1R_4LC_1C_4) = 0 \qquad \text{I}^1$$

and

$$F(\omega R_1R_2R_3R_4LC_4) = 0. \qquad \text{II}^1$$

Elimination of ω shows that a certain relation must exist between the constants of the circuit.

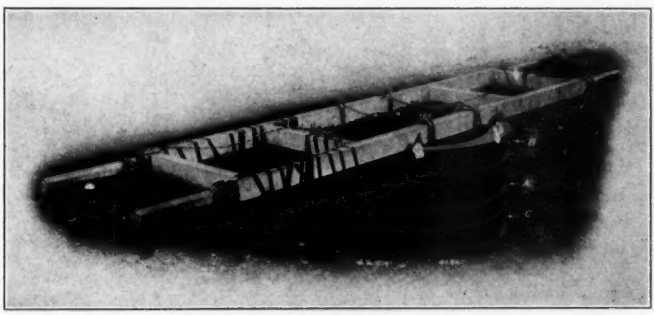


FIGURE 4

The N. A. C. A. bomb detector

If this condition is satisfied the frequency at which balance is possible is given by equation:

$$\omega_0 = \varphi(R_1R_2LC_1C_4)$$

At all other frequencies the bridge is out of balance.

It is obvious that it is very difficult to maintain a pure sine wave of a constant frequency. For this reason a Maxwell bridge is not satisfactory. It is immediately obvious from equation I that if the ratio C_1/C_4 is of the order of, for instance, one per cent, an irregular wave form cannot be balanced out to a satisfactory degree.

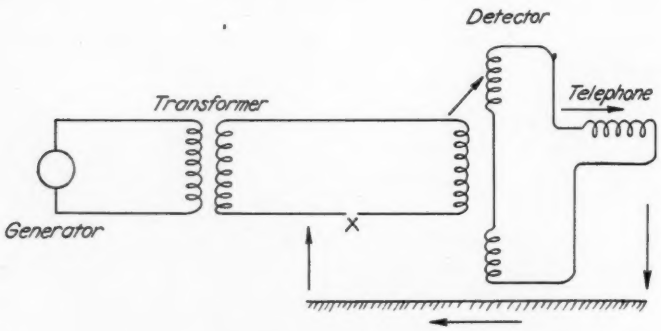
This condition may be further aggravated by the fact that the harmonics are often more audible than the fundamental frequency employed. This shows the nature of the difficulty. The latter bridge acts as a band elimi-

nation filter. The question arises: Is it possible to devise some method by which the bridge current actually is zero instead of some minimum value?

It is well known that the human ear is capable of differentiating between two sounds of intensity, 0 and 1, while the difference in sounds designated by the intensities 4 and 5 can only be detected by a trained operator. The power is, in this latter case, increased by just one Deci-Bel.

It is thus of paramount importance to approach the condition of perfect balance. If the balancing is poor it is perfectly useless to increase the power employed. In fact, it has even been proposed to employ a current of small amplitude.[2]

The author considered the possibility of a scheme by which it would be possible to obtain a more complete balance. It was realized that if this



PATH OF CURRENT THROUGH TELEPHONE WITH ONE POWER LINE DISCONNECTED AT X.
FIGURE 5

could be done the amount of "power" used could be increased accordingly. The N. A. C. A. detector is a result of this study.

It was conceived that a strong alternating magnetic field could be used, the idea being simply to compare the field strengths at two arbitrary points. After going through various stages, the arrangement shown in figure 3 was finally adopted. A is the "power" coil producing the field. The magnitudes compared are, in this case, the flux passing through the two "pick-up" coils *B* and *C*. The sole function of the primary is to produce an alternating field in the buried object. To meet this condition efficiently the diameter of the coil should not be made much smaller than the thickness of the layer to be searched. There is no reason whatsoever why the "power" coil and the "pick-up" coils should be brought close together. In fact, they should be arranged as far apart as the constructional conditions will permit.

The field may be considered to consist of the undisturbed power field and the superimposed field which is induced in the object. It is the function of the "pick-up" system to detect this field issuing from the object. In order to perform this latter function properly, the "pick-up" system must, of course, be brought close to the object, but not, necessarily, close to the primary. The arrangement shown in figure 4 was found to be a desirable practical solution. A is the "power" coil, B and C the two opposing "pick-up" coils.

As pointed out before, the sensitivity of the instrument depends entirely

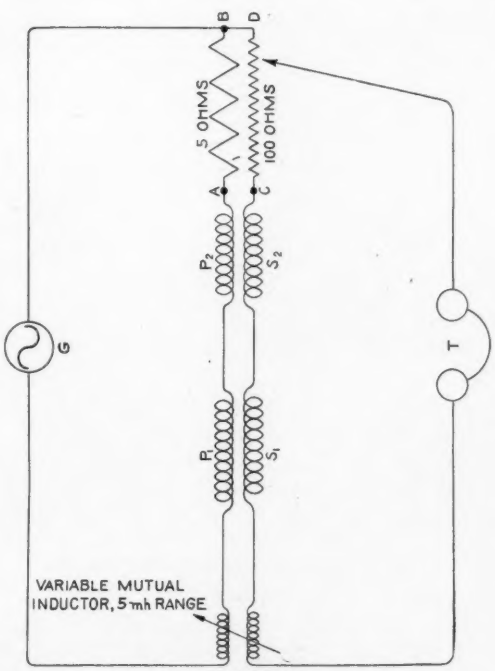


FIGURE 6

on the degree to which balancing is possible. No source of current is entirely free from harmonics, nor is it possible to maintain a given frequency for any length of time. It is essential that the balancing of the detector be independent of the frequency. This requirement has been met quite closely in the N. A. C. A. type. The balance of the detector was actually so perfect that the operator had a tendency to forget the note, and even preferred to keep the instrument slightly out of balance.

As regards the primary coil, A, there is no difficulty. The self-capacity

is of no great avail. One point of some importance is the fact that the self-capacity of the secondaries has been reduced to a minimum by employing single-layer coils. The stray capacity between primary and secondaries has, in particular, been reduced to a minimum by the large spacing between the coils. This is the vital point. The following experiment will clearly indicate the nature of this latter effect. It happened during the field work that one power line was broken or disconnected, as in figure 5. Considerable charging current would then pass through the telephone in the manner indicated, throwing the detector entirely out of balance. The primary to secondary capacity is, consequently, of sufficient magnitude to pass an audible current as soon as the average potential of the "power" coil is fluctuating to any noticeable extent. This current is clearly due to difference in potential between the "power" coil and the observer. This simple experiment shows again the importance of suppressing undesired capacity effects.

In this connection, we wish to point to the limitations of the scheme

FIGURE 7

Detector in operation

disclosed in Drawing No. 5838 of the Bureau of Standards, termed "Induction-balance for the Detection of Metallic Bodies," and dated June 16, 1922. This arrangement, which resembles one described by M. C. Gutton in 1915,[2] employs two transformers in form of coils five feet in diameter, with primary and secondary close together. The system is shown in figure 6.

A considerable difference in potential exists between P_1 and S_1. The higher harmonics of the source pass across at this point, and probably also in the mutual inductor. Approximately one-half of this current goes through the phones. For this reason the instrument cannot be expected to possess any great sensitivity.

Regarding the question of "power" actually required for the operation of the instrument, it will be observed that, since no resistances form part of the circuit, the only energy consumption is theoretically occurring in the telephone membrane. In other words, if the coils were made of wire having no resistance, all of the "power" supplied would appear in the telephone and as heat in the object.

The telephone employed was especially designed to match the low impedance of the "pick-up" system. The current employed for field operation of the instrument was supplied by a small, five-hundred cycle generator.[3] A photograph of the instrument in operation on the site is shown in figure 7. Two men carried the detector with the bottom coil a few inches above the ground. The telephone operator walked behind, or to one side, at a distance of about twenty feet. No difficulty was experienced in operation. The instrument showed no tendency to get out of balance, nor was the balancing more difficult than the tuning of a single-dial radio. The in-

FIGURE 8

strument was adjusted by changing the position of the "power" coil or, in fact, by touching slightly one of its turns, which acted as a very good vernier. The stability of the instrument was, however, so perfect that on starting out for field work, new adjustment was not always needed.

Several bombs were located after the first day's work, one of which was a live T.N.T. seventeen-pounder, buried with its center of gravity approximately two feet below the surface. While the destroyed bombs all were found in holes of considerable size, no visible evidence disclosed the presence of the live bombs found during the search.

[1] "War Works," The Bureau of Standards.

[2] *Compt. Rend.*, **161** 1915, page 171, by M. C. Gutton.

[3] For further data see the September issue of *J. Franklin Institute.*

THE HEAT OF FORMATION OF WATER[1,2]

BY FREDERICK D. ROSSINI

NATIONAL BUREAU OF STANDARDS, WASHINGTON, D. C.

Read before the Academy September 19, 1930

As part of the program of thermochemical research now being carried on at the National Bureau of Standards, the heat of formation of water has been redetermined. The importance of this thermal quantity lies in its direct use in the calculation of the heat of formation of practically all organic and of many inorganic compounds, and in its indirect use in thermodynamic calculations for chemical reactions involving H_2O.

During the past century there have appeared fourteen papers reporting measurements of the heat of formation of water. Up to the present time the usually accepted value[3] has been based upon the measurements of Thomsen[4] in 1873, Schuller and Wartha[5] in 1877, and Mixter[6] in 1903. Of these, the data of Schuller and Wartha are the most reliable, yet even their average value has attached to it an uncertainty of some 8 parts per 10,000.

The method employed in this investigation was as follows: The apparatus consisted of a calorimeter surrounded by a constant temperature jacket. The calorimeter was composed of a calorimeter can containing a measured mass of water, a device for measuring changes of temperature, a stirring mechanism, a reaction vessel for burning the gases at constant pressure and a heating coil. The problem was to determine the quantitative correspondence between the thermal energy liberated (1) by the reaction between hydrogen and oxygen to form a measured mass of water and (2) by a measured quantity of electrical energy, using the calorimeter system as the absorber of the two quantities of energy and its temperature rise as the comparator. A combustion experiment, in which m grams of water were formed in the reaction vessel, produced a temperature rise of Δt degrees in the calorimeter at some average temperature t_1. In a calibration experiment, the same temperature rise of Δt degrees, at the same average temperature t_1, was brought about in the same calorimeter by a measured quantity of electrical energy, E joules. In this manner, the formation of m grams of water was found to be equivalent, in energy evolved, to E joules of electrical energy.

The supreme advantage of this method lay in the fact that the accuracy of the value obtained for the heat of formation of water depended only upon the determination of the mass of water formed and of the quantity of electrical energy. This required accurate standards of mass, and of the ohm, the volt and the second, all of which were available at this bureau. High precision was obtained by the use of proper calorimetric technic,

a sensitive platinum resistance thermometer for measuring changes of temperature, a sensitive potentiometric system for measuring the electrical power input, a precision timing device and a suitable balance for determining the mass of water formed.

Figure 1 shows the assembly of the apparatus. *A*, *B*, *C* and *D* comprise the gas purifying system; *E*, *F* are flow meters; *G* shows a group of stopcocks; *J*, *H* are inlet tubes; *K* is the reaction vessel and support; *L*, the

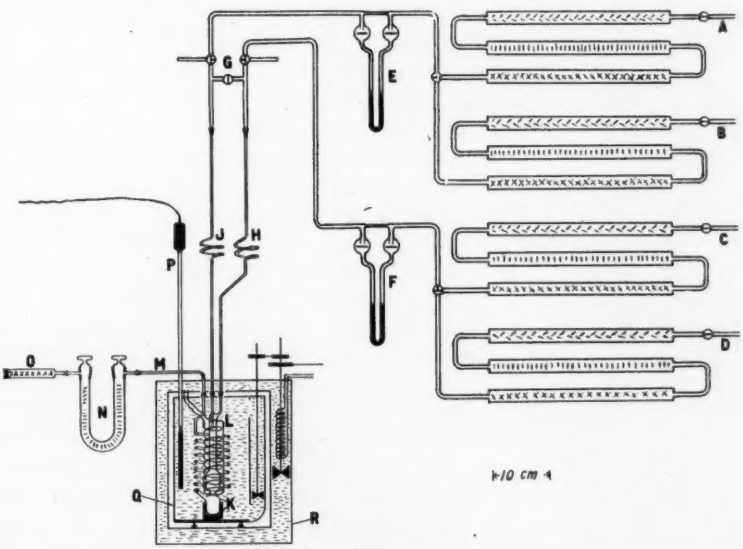


FIGURE 1

Schematic diagram of the apparatus. A, B, C, D, purifying tubes; E, F, flow meters; G, stopcocks; J, H, inlet tubes; K, reaction vessel and support; L, calorimeter heater; M, connecting tube; N, U-tube absorber; O, guard tube; P, platinum thermometer; Q, calorimeter can; R, jacket.

calorimetric heater; *M*, the connecting tube; *N*, the U-tube absorber in which the H_2O formed is collected and weighed; *O*, the guard tube; *P*, the platinum resistance thermometer; *Q*, the calorimeter can; *R*, the calorimeter jacket.

Figure 2 shows a cross section of the reaction vessel and support. At *A* are the spark leads; *B*, *C* designate the gas inlet tubes; *D*, the exit tube; *E*, the cooling coil; *F*, the burner tube; *G*, the reaction chamber; *H*, the condensing chamber; *J*, the supporting frame. Except for the end of the burner tube, which is silica glass, the reaction vessel is made of Pyrex laboratory glass. In an actual combustion the flame, ignited by means of

sparks jumping across the gap at *F*, burned quietly at the tip, and the condensed H_2O collected in the chamber below. In figure 3 are curves showing the variation of the calorimeter temperature with time for combustion experiment S–C and calibration experiment S-1. The ordinate scale gives the temperature in °C. while the abscissa scale gives the time in minutes. These curves portray the substantial identity of the calorimetric conditions in the calibration and in the combustion experiments.

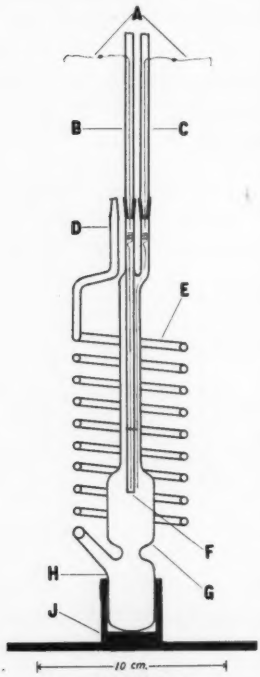


FIGURE 2

Cross-section of the reaction vessel and supporting frame. A, leads of spark circuit; B, C, inlet tubes; D, exit tube; E, cooling coil; F, burner tube; G, reaction chamber; H, condensing chamber; J, supporting frame.

The amount of reaction in each combustion experiment was determined from the mass of H_2O formed, not from the mass of hydrogen or of oxygen consumed. The presence of inert and nonreactive impurities in the hydrogen or oxygen, therefore, would not affect the experiments chemically or calorimetrically. This being the case, an examination of the reaction products gave definite and complete evidence as to the purity of the reaction.

The hydrogen and oxygen were regular commercial stock, the former of electrolytic and the latter of atmospheric origin. Analysis of the hydrogen and oxygen after passage through the purifying system was made by the Gas Chemistry section of this bureau. The amount of inert gas in the oxygen was found to be, in two analyses, 0.63% and 0.52%. About two-thirds of this was nitrogen and the remainder argon plus neon. The amount of gas in the hydrogen which was non-combustible with oxygen was found to be, in two analyses, 0.09% and 0.06%. This impurity was probably oxygen.

No carbon dioxide, sulphur dioxide or nitrogen oxides were found in the reaction products. The amount of NH_3 formed when oxygen was burned in hydrogen was found to be 1 mole NH_3 per 250,000 moles of H_2O. This introduces a correction to the heat effect of about 1 part in 300,000.

The amount of H_2O formed in each combustion experiment was determined, after the calorimetric observations were completed by sweeping out all the H_2O in the reaction vessel with oxygen or hydrogen and absorbing it in a weighed glass-stoppered U-tube containing "Dehydrite" ($Mg(ClO_4)_2.3H_2O$) and P_2O_5. The increase in weight of the absorber,

about 2.85 grams, was determined to ±0.05 mg. and corrected to vacuum to give the true mass of H_2O formed.

For the energy measurements, the potential drops across the standard resistance and the calorimeter heating coil were measured on a White double potentiometer, and the time was measured with a device similar to that of Johnson,[7] but possessing several improvements suggested by him. The unit of energy in this work is the international joule defined in terms of the mean solar second and standards of the international ohm and the international volt as maintained at this Bureau.

The main temperature rise of about 2.8°C occurred in 12 minutes, and was measured with a platinum resistance thermometer having in its measuring system a Mueller thermometer bridge with a galvanometer eight meters distant from a scale on which 1 mm. was approximately equivalent to 0.0002° C.

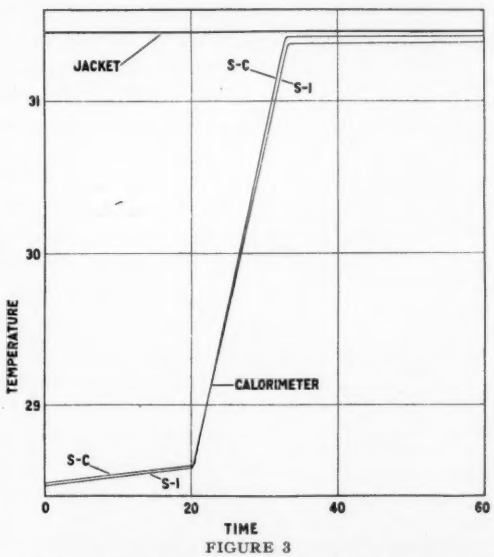


FIGURE 3

Plot showing the variation of the calorimeter temperature with time. The ordinate scale gives the temperature in °C. and the abscissa scale marks the time in minutes. Curve S-1 is for electric calibration experiment S-1, and curve S-C is for combustion experiment S-C.

Three sources or sinks of energy existed in the combustion experiment which were not present in the calibration experiments: (1) the "spark" energy given to the calorimeter system in the sparking operation, (2) the "gas" energy required to bring the entering gases to the temperature of the calorimeter, and (3) the "vaporization" energy involved in the formation of a small amount of H_2O as vapor. The "spark" energy was determined experimentally while the "gas" energy and the "vaporization" energy were calculated.

Two sets of experiments were made. Set I consisted of 11 calibration and 9 combustion experiments at 25°C., and Set II of 5 calibration and 9 combustion experiments at 30°C. The latter values were corrected to 25°C. for comparison with Set I. In some experiments, hydrogen was

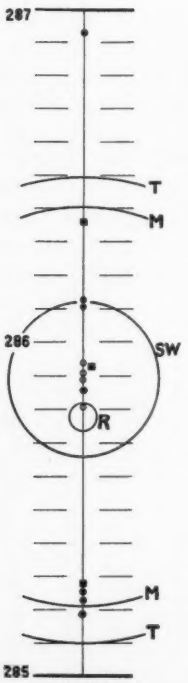


FIGURE 4

Plot of the data on the heat of formation of water.

The ordinate scale is in international kilojoules per mole and the data is computed to give the heat of formation of liquid H_2O at 25°C., 1 atm. The points are: ○, Schuller and Wartha; ■, Thomsen; ●, Mixter. The small circle shows the assigned limit of error of the value obtained in the present work, while the other circle and arcs show the assigned errors for the recomputed data.

burned in oxygen, and in others, oxygen in hydrogen. The value obtained for the heat of formation of 1 mole (18.0156 g.) of liquid H_2O at 25°C. and a constant pressure of 1 atm., is 285,775 ± 40 international joules.

The data of Schuller and Wartha, Thomsen and Mixter have been recomputed in terms of the international joule as the unit of energy. Schuller and Wartha calibrated their ice calorimeter in terms of the mean calorie, whose value in terms of the international joule has just been determined at this Bureau.[8] Both Thomsen and Mixter calculated the heat capacity of their calorimeter system from that of the parts.

In figure 4 are plotted the recomputed values of the individual experiments of these investigators. The four experiments of Schuller and Wartha are indicated by small circles, the three of Thomsen by black squares, and the seven of Mixter by solid black circles. The radii of the large circles and arcs measure in each case the assigned error.

Figure 5 shows a magnified view of that portion of figure 4 in which the individual values of the present experiments fall. The experiments in which oxygen was burned in hydrogen are indicated with an appendage. The heavy black horizontal lines show the average values of the recomputed data: Schuller and Wartha, 285,890 ± 235; Mixter, 285,810 ± 600; Thomsen, 285,820 ± 700.

The value obtained in this investigation for the heat of formation of one mole (18.0156 g.) of liquid H_2O from gaseous hydrogen and oxygen at 25° C. and a constant pressure of 1 atm., is 285,775 ± 40 international joules. Using the factor 1.0004 and $\frac{1}{4.185}$ for converting, respectively, international joules to absolute joules[9] and absolute joules to g.-cal.$_{15}$,[10] this value becomes 285,890 ± 40 absolute joules or 68,313 ± 10 g.-cal.$_{15}$.

The author is greatly indebted to E. W. Washburn, under whose direction this work was executed, for his deep interest and valued suggestions. Many thanks are due E. F. Mueller for his advice on the

calorimetric technic, to F. A. Smith for his aid in designing the reaction vessel and to the various analytical, calibration and standardization sections of this Bureau which contributed assistance.

[1] Publication approved by the Director of the Bureau of Standards, U. S. Department of Commerce.

[2] Full details of this work will appear in the Bureau of Standards *Journal of Research.*

[3] (*a*) Lewis, *J. Am. Chem. Soc.*, **28,** p. 1389, 1906; (*b*) Roth, *Zeitst. Electrochemie,* **26,** p. 288, 1920.

[4] Thomsen, *Ann. Physik und Chemie,* **148,** p. 368, 1873.

[5] Schuller and Wartha, *Ann. Physik und Chemie,* **2,** p. 359, 1877.

[6] Mixter, *Am. J. Sci.*, **16,** p. 214, 1903.

[7] Johnston, *J. Opt. Soc. Am.* and *Rev. Sci. Instr.*, **17,** p. 381, 1928.

[8] Osborne, Stimson and Fiock, *Bur. Stand. J. Res.*, **5,** 501, 502, 1930.

[9] *Tech. News Bull. Bur. Stand.*, No. 156, April, 1930.

[10] *International Critical Tables,* **1,** p. 24.

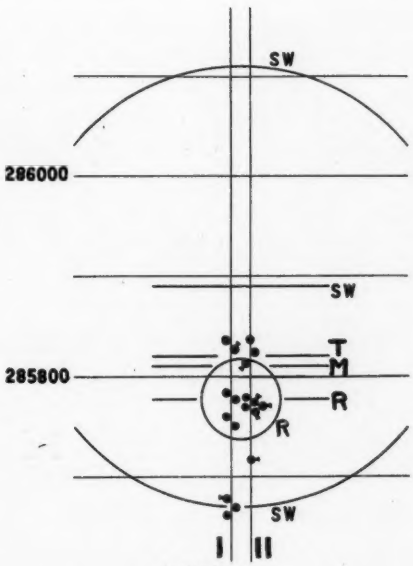


FIGURE 5

Plot of the present data on the heat of formation of water.

The ordinate scale gives the heat of formation of liquid H_2O in international joules per mole at 25°C., 1 atm. The O_2 in H_2 experiments are denoted with an appendage. The experiments of Set I are distinguished from those of Set II. The average values of Schuller and Wartha, Mixter, Thomsen, and of the present data are shown, together with the "error" circles for the data of Schuller and Wartha and the present work.

CRYSTALLINE ALPHA AND BETA METHYL-d-GULOSIDES

BY HORACE S. ISBELL

NATIONAL BUREAU OF STANDARDS, WASHINGTON, D. C.

Read before the Academy September 23, 1930

In 1893 Emil Fischer discovered that two isomeric methyl glucosides are formed from glucose by the action of methyl alcohol in the presence of hydrochloric acid. Subsequently, Fischer's method has been applied to other sugars giving many similar compounds. The reaction yields a mixture of glucosides from which, frequently, crystalline products have not

been obtained. These mixtures are very complex, probably containing many isomers, the separation of which is extremely desirable. Since in the aldohexose series only methyl glucosides of glucose, galactose and mannose were known in the crystalline state, the preparation of additional methyl glucosides was undertaken.

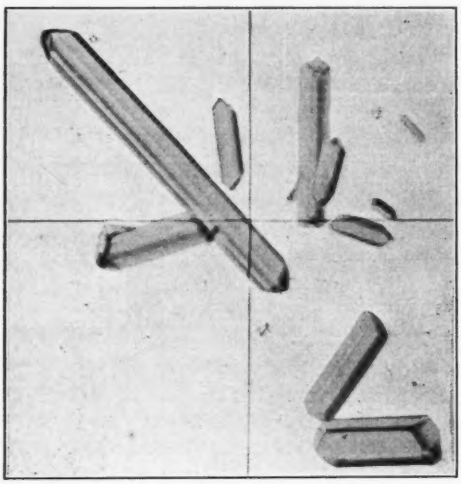


FIGURE 1

α-Methyl-*d*-guloside $CaCl_2 \cdot 2H_2O$.

Recently the writer found a new calcium chloride compound of *d*-gulose which makes available a crystalline form of that sugar. The new compound may be used as a raw material for the preparation of other derivatives. The synthesis of the methyl gulosides was first attempted by Blanksma and Van Ekenstein who obtained a sirup which they were unable to bring to crystallization. Their work was repeated, starting with pure gulose which was prepared by removing the calcium chloride from α-*d*-gulose $CaCl_2 \cdot H_2O$ with silver oxalate. A relatively pure product was obtained but it could not be brought to crystallization, and after a thorough trial, it was evident that a new method was necessary to reach the desired substance. Although crystalline calcium chloride compounds of the glyco-sides were unknown, the encouraging results obtained by the application of the calcium chloride compounds in the preparation of sugars which had not been previously crystallized, led the writer to apply the same principle in the separation of the methyl gulosides. This resulted in the preparation of crystalline calcium chloride compounds of α- and β-methyl-*d*-gulosides. These compounds served for the separation of the α- and β-isomers from the complex reaction product and, ultimately, gave the pure methyl gulosides in the crystalline state.

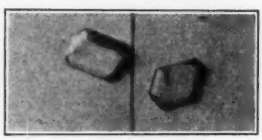


FIGURE 2

$(\alpha$-Methyl-*d*-guloside$)_2$ $CaCl_2 \cdot 3H_2O$.

The new substances were first prepared by the addition of calcium chlo-

ride to the sirup obtained by the method of Blanksma and Van Ekenstein. Subsequently, the method was simplified by preparing the calcium chloride gulosides direct from *d*-gulose calcium chloride in the following manner: The calcium chloride compound was refluxed with methyl alcohol containing about 1 per cent hydrogen chloride. After several hours an excess of powdered calcium carbonate was added; the hydrochloric acid was neutralized, forming calcium chloride. This additional salt did not interfere with the crystallization of the calcium chloride methyl gulosides which were easily obtained upon evaporation of the solution. The first crystals separated in slender needle-like prisms which gave a specific rotation of about +60°. In the preliminary work the calcium chloride was removed from this crude product by means of silver oxalate. After considerable difficulty, a crystalline product was obtained from the resulting solution which, after drying in vacuo over P_2O_5, gave the correct analysis for methyl guloside. Its specific rotation, 106°, was commensurate with the value derived from the calcium

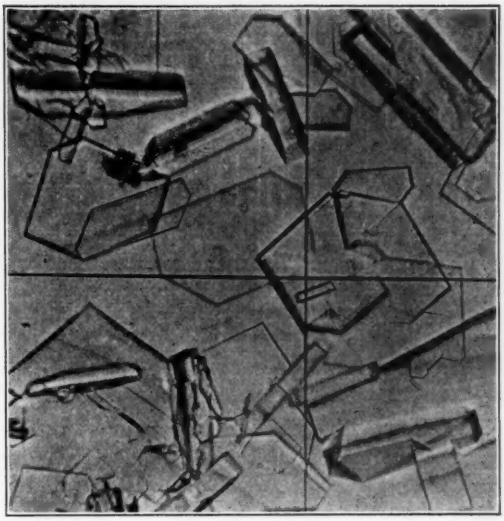


FIGURE 3
α-Methyl-*d*-guloside·H_2O.

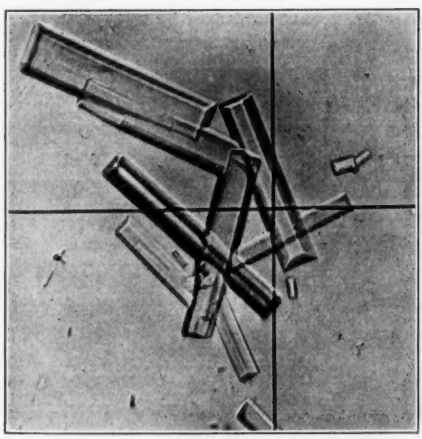


FIGURE 4
β-Methyl-*d*-guloside $CaCl_2 \cdot 2H_2O$.

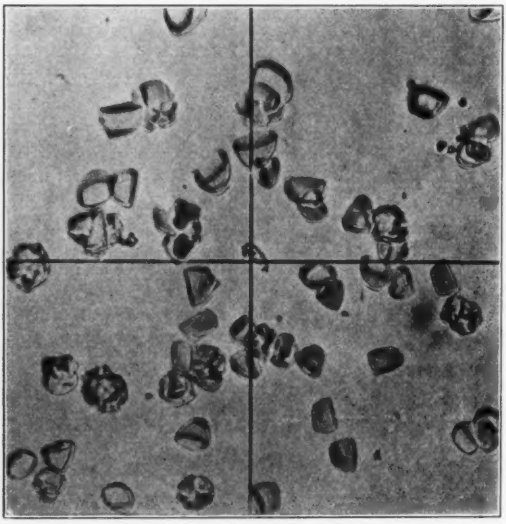


FIGURE 5

(β-Methyl-*d*-guloside)$_2$ $CaCl_2$.

chloride compound from which it was prepared and the molecular rotation (20,600) differed normally from the molecular rotation (−16,-100) of β-methyl-*d*-guloside. However, subsequent work showed that the appearance of the crude calcium chloride compound changed on recrystallization. The crude product crystallizes in faintly double refracting needles while the recrystallized product separates in strongly double refracting truncated prisms (Figs. 1 and 2). After many crystallizations a product was obtained, the analysis of which corresponded to (α-methyl-*d*-guloside)$_2$, $CaCl_2.3H_2O$. The specific rotation is +83 which corresponds to a somewhat higher molecular rotation than the rotation of the product which was first obtained. After removing the calcium chloride the corresponding guloside was readily obtained as the monohydrate (Fig. 3). It decomposed at 77° losing its water of crystallization. The specific rotation ($[\alpha]_D^{20} = 109$) agrees with the value from the

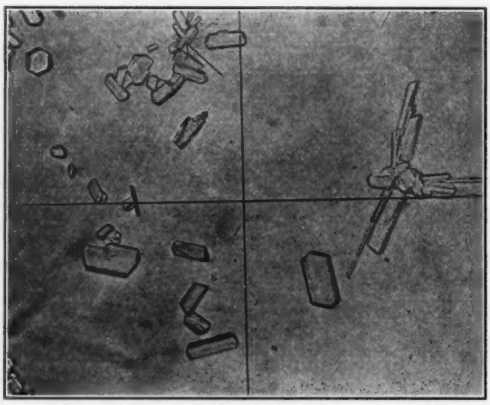


FIGURE 6

β-Methyl-*d*-guloside.

recrystallized calcium chloride compound. The molecular rotation (23,100) differs from the molecular rotation of β-methyl-*d*-guloside (−16,100) by slightly more (39,200) than the normal value (37,000). The question remained whether the calcium chloride compounds which was first obtained constituted a different form of the guloside or whether it contained some impurity which lowered its rotation. This was answered by the preparation of both calcium chloride compounds from pure α-methyl-*d*-guloside (monohydrate). The first compound was obtained by adding an excess of calcium chloride, while the second compound was obtained by using the calculated amount of the salt. The specific rotation of the first compound when prepared from the pure guloside is +67°. This is equivalent to a molecular rotation of 22,900 which is in accord with the values 23,100 and 23,000 obtained from α-methyl-*d*-guloside monohydrate and (α-methyl-*d*-guloside), $CaCl_2.3H_2O$, respectively.

The mother liquor from the preparation of the calcium chloride compounds just described gave, on standing, a third calcium chloride methyl guloside. The substance separated in rectangular prisms which were moderately soluble in ethyl alcohol. The analysis of the crude product corresponded to the formula β-methyl-*d*-guloside $CaCl_2.2H_2O$. It melted at 155° and gave a specific rotation of −46°. Upon recrystallization from absolute ethyl alcohol this compound changed into a fourth calcium chloride methyl guloside, which was easily obtained pure because it is very insoluble in absolute ethyl alcohol (Figs. 4 and 5). It gave a specific rotation of −65 and the analysis corresponds to the formula (β-methyl-*d*-guloside)$_2$ $CaCl_2$. After removing the calcium chloride with silver oxalate β-methyl-*d*-guloside was easily obtained in the crystalline state (Fig. 6). It melts at 176° (uncorrected) and gives a specific rotation of −83°. This value is equivalent to a molecular rotation of 16,100 which is in agreement with the equivalent rotation 16,200 of the parent calcium chloride compound.

The method for the separation of the methyl gulosides which has just been enumerated may be applicable for the separation of other glycosides as yet unknown in the crystalline state. It is not to be expected that glycosides of all sugars will form crystalline compounds with calcium chloride or other salts. However, a number of compounds of this type have already been prepared and undoubtedly many more will be found in the future.

THE RING STRUCTURE OF MANNOSE

BY HORACE S. ISBELL

NATIONAL BUREAU OF STANDARDS, WASHINGTON, D. C.

Read before the Academy September 23, 1930

For a long time the structure of the sugars has been the subject of considerable controversy. Originally the sugars were considered to be straight chain aldehydes, but they are now believed to have a cyclic or ring structure. The size of the ring may be determined experimentally by methylating the free hydroxyls and determining where the groups are located by degradation. Another method for determining ring structure has been devised by C. S. Hudson. This method depends upon the comparison of optical rotations. Those substances whose rotations conform to certain rules which he has developed are considered to have the same ring structure, while those which do not agree are considered to have different ring structures. Although there are other instances where the results from the two methods do not agree, the discussion today will be restricted to the ring structure of mannose.

The sugar mannose differs from glucose in that the hydroxyl on the second carbon atom lies to the left rather than to the right as in glucose. The difference in the molecular rotations of like forms of glucose and mannose is presumably twice the rotation of the second carbon atom, which is

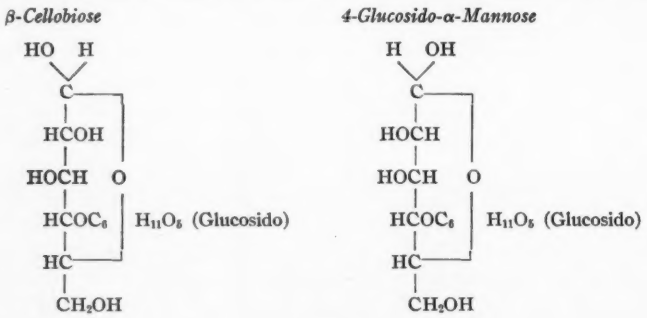


FIGURE 1

termed the epimeric difference in rotation. The value for this difference obtained from α-mannose and α-glucose is 14,900, while a value of 6500 is obtained from β-mannose and β-glucose. According to the van't Hoff theory of optical superposition the same value should be obtained from the two pairs. The question is, what difference in the structures of the compounds causes the different values. According to Hudson, the substances

differ in ring structure. He believes that α-mannose and α-methyl mannoside have the 1,4 ring, while β-mannose, α- and β-glucose and α- and β-methyl glucosides have the 1,5 ring. The 1,4 ring for α-methyl mannoside is not in agreement with the results obtained from methylation studies by Haworth and other workers. In order to decide between the two views Hudson has recently devised a very ingenious test. He states that "a decision between the rival ring classifications might be obtainable from measurements of the rotations of a pair of substances having configurations in which only one of the 1,4 and 1,5- rings could be assumed to exist." He pointed out that cellobiose and 4-glucosido-mannose are a pair of substances which fulfill these requirements.

In figure 1 the structure of 4-glucosido-mannose and cellobiose are shown. It is seen that in each a glucosido group is substituted for the hydroxyl group on the fourth carbon atom; consequently the compounds cannot form a 1,4 ring.

Hudson found that the difference in the molecular rotations of the halogen acetyl derivatives of these two sugars (11,800) agreed with the value (11,300) which he obtained by comparison of certain acetylated derivatives of glucose and mannose. But it did not agree with the value obtained from the acetylated α-methyl glucoside and mannoside. This, he believed, proved that the values obtained according to his classification are correct and that, since the value obtained from the acetylated α-methyl mannoside and glucoside is not confirmed, there is no escape from the conclusion that Haworth's assignment of the same ring to normal methyl glucoside and mannoside must be invalid.

The importance of this conclusion demands that the evidence be conclusive. A search of the data in the literature shows that 4-glucosido-mannose has been prepared, and although its initial rotation was not reported, the rotation of the substance 7 minutes after solution indicated that it did not support Hudson's deduction. In the light of this the writer decided to repeat the preparation of 4-glucosido-mannose and to attempt the preparation of a sufficient number of derivatives of the sugar to determine whether the conclusion, that α-mannose and α-methyl mannoside have the 1,4 ring structure, is supported by the rotations of all the derivatives of 4-glucosido-mannose.

In 1921 Bergmann and Schotte first prepared 4-glucosido-mannose monohydrate by the oxidation of cellobial. Subsequently the same compound was prepared by Brauns by means of the action of anhydrous hydrogen fluoride upon cellobiose. Bergmann and Schotte gave the specific rotation of the substance 7 minutes after solution as +15.07, becoming constant at 10.65, while Brauns gave +7.3, 15 minutes after solution, becoming constant at 5.8. The writer has repeated Dr. Brauns' work. The initial rotation of the monohydrate of the sugar extrapolated to zero time was

found to be +14.6, with the final rotation of 5.88 in agreement with Dr. Brauns' value.

It is now possible to compare the molecular rotation of β-cellobiose and 4-glucosido-α-mannose. The difference in the molecular rotations of these substances is equal to

$$2R_2 - 2A_{OH} = 212.$$

The value of $2A_{OH}$ has been determined by Hudson and found to be about 16,700, which gives for $2R_2$ the epimeric difference 16,900. This value differs widely from the value (6700) obtained according to Hudson's classification of ring structures. However, it agrees more closely with the values (14,900 and 15,300) obtained from α-mannose and α-methyl-mannoside. Since 4-glucosido-mannose cannot form a 1,4 ring, this parallelism with α-mannose is strong evidence that neither α-mannose nor α-methyl-mannoside has the 1,4 ring structure.

TABLE 1

EPIMERIC DIFFERENCES IN MOLECULAR ROTATIONS FOR THE MANNOSE SERIES

	$[\alpha]_D$	[M]	EPIMERIC DIFFERENCE $2R_2$
Alpha Forms			
α-Glucose	+113	20,340	14,940
α-Mannose	+ 30	5,400	
α-Methyl glucoside	+157.9	30,630	15,300
α-Methyl mannoside	+ 79	15,330	
β-Cellobiose	+ 16	5,470	16,900 (212 + $2A_{OH}$)
4-Glucosido-α-mannose	+ 14.6	5,258	
Beta Forms			
β-Glucose	+ 19	3,420	6,480
β-Mannose	− 17	− 3,060	
β-Cellobiose	+ 16	5,470	7,180
4-Glucosido-β-mannose	− 5	− 1,710	

It has just been shown that the α-forms of a pair of sugars suggested by Hudson as suitable to decide between the rival ring classifications give an epimeric difference which is in agreement with Haworth's classification of normal methyl glucoside and mannoside.

We will now consider the beta forms of the same disaccharides in order to establish a parallelism with the rotations of beta mannose. 4-Glucosido-β-mannose was not reported in the literature; accordingly it was necessary to prepare it for the first time. It was obtained by crystallization from methyl alcohol. The first product was mixed with the crystalline anhydrous alpha form which was also obtained for the first time. The separation of the two substances was difficult but the beta compound was finally obtained in what appeared to be the pure state. It separates in slender prisms which melt at 205° and give a specific rotation of about −5°.

The difference in the molecular rotations of cellobiose and 4-glucosido-β-mannose is approximately 7200. This value is in agreement with Hudson's classification, or in other words, with the value obtained from β-mannose and normal glucose. Thus it appears that 4-glucosido-α-mannose is a substituted α-mannose, while 4-glucosido-β-mannose is a substituted β-mannose. The rotations of the alpha forms give an epimeric difference of about 16,900, while those of the beta give a difference of about 7200.

The parallelism between the optical rotations of the α- and β-forms of 4-glucosido-mannose with the corresponding forms of mannose indicates that neither α- nor β-mannose has a 1,4 ring structure. We cannot definitely state at this time which form of mannose has a structure similar to the normal form of glucose. It is entirely possible that Hudson's classification of ring structures is in reality a classification of a new type of isomerism. It seems reasonable to presume that there is a difference in the structures of α- and β-mannose, the exact nature of which must remain for the future to determine.

The writer expresses his appreciation to Dr. D. H. Brauns for the loan of the apparatus with which he originally prepared 4-glucosido-α-mannose. Appreciation is also expressed to other members of the Polarimetry Section who have aided in the course of the investigation.

NON-DISJUNCTION OF THE X-CHROMOSOME IN DROSOPHILA VIRILIS[1]

BY M. DEMEREC AND J. G. FARROW

DEPARTMENT OF GENETICS, CARNEGIE INSTITUTION OF WASHINGTON, COLD SPRING HARBOR, N. Y.

Communicated October 14, 1930

An extensive series of experiments were made to test the effect of x-ray radiation on the mutability of the mutable reddish-alpha and the miniature-beta genes. These experiments were planned in such a way as to give results which could be used to determine the effect of x-rays on the mutable genes and also to study non-disjunction of the X-chromosome. The results related to the problem mentioned last will be reported in this paper.

PRIMARY NON-DISJUNCTIONS. *Frequency.*—Data on the frequency of primary non-disjunctions were obtained by breeding F_1 females from crosses involving several sex-linked characters. That precaution made it certain that these females were not non-disjunctional and to a large extent eliminated the possibility of inclusion in the results of secondary non-disjunctions. Since, however, there was practically no way of making certain that the male parents of the F_1 females were not XYY, it cannot be claimed

TABLE 1

DATA SHOWING THE NUMBER OF REGULAR AND NON-DISJUNCTIONAL OFFSPRING

TREATMENT	MATING	NUMBER OF EXPERIMENTS	NUMBER OF CULTURES	REGULAR OFFSPRING ♀	REGULAR OFFSPRING ♂	REGULAR OFFSPRING TOTAL	NON-DISJUNCTIONS ♀	NON-DISJUNCTIONS ♂	NON-DISJUNCTIONS TOTAL	PER CENT NON-DISJUNCTIONS ♂	PER CENT NON-DISJUNCTIONS TOTAL
					Primary Non-Disjunctions						
None	*se y sc v/re-α*x *wh*	2	143	7,291	6,181	13,472	1	9	10	0.145	0.074
None	*v mt-1 sb/mt-β*x *wh*	1	118	9,607	6,190	15,797	—	3	3	0.048	0.019
None	*mt-1 f/mt-β*x *wh*	1	48	3,973	3,946	7,919	1	11	12	0.279	0.152
Total		4	309	20,871	16,317	37,188	2	23	25	0.141	0.067
Females x-rayed	*se y sc v/re-α*x *wh*	4	167	7,292	6,299	13,591	—	68	68		
Females x-rayed	*v mt-1 sb/mt-β*x *wh*	1	185	14,969	12,986	27,955	5	149	154		
Females x-rayed	*mt-1 f/mt-β*x *wh*	1	189	12,588	11,675	24,263	12	121	133		
Total		6	541	34,849	30,960	65,809	17	338	355		
Grand total		10	850	55,720	47,277	102,997	19	361	380		
					Secondary Non-Disjunctions						
Our data				1,848	1,606	2,554	9	8	17		0.666
Weinstein's data[2]				792	700	1,492	8	11	19		1.273
Total				2,640	2,306	4,046	17	19	36		0.889

with certainty that all of the F_1 females used in the experiments were not XXY females. However, sterility of all non-disjunctional males tested the low proportion of non-disjunctional females, and the frequent occurrence of equational non-disjunctions makes it very probable that at least a great majority of the F_1 females were regular XX females, and that the non-disjunctions observed were primary non-disjunctions.

The frequency of the primary non-disjunctions is shown in the first section of table 1 where data on the offspring of the non-treated females are given. In a total of 37,188 flies, 25 exceptions were found which amounts to one exception in about 1500 flies.

Female to Male Ratio.—As it has been found previously in *Drosophila melanogaster*, in the present case also, exceptional females were fewer than exceptional males. The female to male ratio in the material investigated was 1:11.5. In the much larger number of individuals obtained by the inclusion of the offspring of the x-rayed females that ratio was 1:19. The difference between these two ratios is probably due to the difference in the number of individuals rather than to the effect of the x-ray radiation. Apparently the variability in the female to male ratio is great as indicated by the differences observed in the ratios obtained in various experiments.

Sterility of Primary Males.—A total of 52 primary males were mated with two virgin females each. None produced any offspring.

Equational Females.—Eight equational females were found. One was a *mt sb/(cv?) mt sb*, two were *mt/mt sb*, and one was *mt/cv mt* from *cv mt sb/mt* parents; and four were *mt f/mt f* from *mt f/mt* parents. From the data available it is not possible to determine the spindle fiber attachment end of the sex-chromosome since an insufficient number of sex-linked characters were involved in the experiments.

SECONDARY NON-DISJUNCTIONS. As it can be seen from the data of the last section of table 1, 36 secondary non-disjunctions were found among 4046 flies. That makes one exception in 1120 flies. The female to male ratio for secondary non-disjunctions was found to be very close to 1 to 1.

DISCUSSION. A primary female is formed when an exceptional XX egg is fertilized by a Y sperm, and a primary male is formed when an exceptional O egg is fertilized by an X sperm. Both the XX egg and the O egg are the two products of the same process, viz., the XX egg is formed, when the two X-chromosomes do not separate at the reduction division; the O egg is formed, when an X-chromosome does not reach the pole at the reduction division. The deficiency of non-disjunctional females indicates that fewer XX than O eggs were formed which is explained by Bridges[3] by an assumption that, frequently, when an abnormal reductional division occurs, one of the X-chromosomes lags on the spindle and does not get included in the nucleus of either of the two newly formed cells (the egg and the polar body). Such an elimination of a chromosome from either of the

two new nuclei would result in a preponderance of O eggs and in a deficiency of XX eggs.

Non-disjunction was investigated in three species of *Drosophila*, viz., in *D. melanogaster*, *D. similans* and *D. virilis*. Female to male ratio for primary non-disjunctions was found to be 1:4.8 in *melanogaster*,[4] 1:13.4 in *similans*[5] and 1:19 in *virilis*. If the deficiency of females were due to the elimination of one of the X-chromosomes at the reduction division, that elimination is apparently more frequent in *virilis* than either in *similans* or in *melanogaster*.

The normal frequency of primary non-disjunctions is slightly higher in *virilis* than it is in *melanogaster*,[4] the frequency being one in 1500 and one in 2000 individuals, respectively. The frequency of abnormal divisions which cause non-disjunctions, however, can be better estimated from the number of male exceptions rather than from the total number of the female and male exceptions. The frequency of the male exceptions in *virilis* was one in 700 and in *melanogaster* one in 1200, indicating that the abnormal reductional divisions involving X-chromosomes are almost twice as frequent in *virilis* females than they are in *melanogaster* females.

As already mentioned, the elimination of an X-chromosome during reduction division is more frequent in *virilis* than it is in *melanogaster*. In *virilis* in 91.3 per cent of abnormal reduction divisions one chromosome is lost during the cell division and in *melanogaster* that loss occurs in 76.4 per cent of cases.

Secondary non-disjunctions are about five times as frequent in *melanogaster* as they are in *virilis*, the frequency being one in 22.5 and one in 112.4 individuals, respectively. This indicates that heterosynapsis in *virilis* is less frequent than heterosynapsis in *melanogaster*. Computed by Bridges'[3] formula the per cent of heterosynapsis in *virilis* is 3.52 and in *melanogaster* 16.5. If the synapsis were to occur with equal frequency between any two chromosomes of an XXY female heterosynapsis would be expected to occur in 67 per cent of all cases.

SUMMARY.—It has been found that: (1) the frequency of primary non-disjunctions was one in about 1500 individuals; (2) the female to male ratio for primary non-disjunctions was 1:19; (3) primary males were sterile; (4) the frequency of secondary non-disjunctions was one in 1120 individuals and (5) the female to male ratio for secondary non-disjunctions was approximately one to one.

[1] This investigation has been aided by a grant from the NATIONAL RESEARCH COUNCIL, Committee on the Effects of Radiation upon Living Organisms.

[2] Weinstein, A., *Sigma XI Quarterly*, **10**, 1922 (45–53).

[3] Bridges, C. B., *Genetics*, **1**, 1916 (1–52, 107–163).

[4] Morgan, T. H., Bridges, C. B., Sturtevant, A. H., *Bibl. Genetica*, **2**, 1925 (1–262).

[5] Sturtevant, A. H., *Zeits. wis. Zoölogie*, **135**, 1929 (323–356).

RELATION BETWEEN THE X-RAY DOSAGE AND THE FREQUENCY OF PRIMARY NON-DISJUNCTIONS OF X-CHROMOSOMES IN DROSOPHILA VIRILIS[1]

BY M. DEMEREC AND J. G. FARROW

DEPARTMENT OF GENETICS, CARNEGIE INSTITUTION OF WASHINGTON, COLD SPRING HARBOR, N. Y.

Communicated October 14, 1930

While studying the effect of x-ray radiation on the mutability of the mutable reddish-alpha and the miniature-beta genes, a close relationship has been noticed between the x-ray dosage applied and the occurrence of primary non-disjunctions. This observation led to a series of experiments which shall be described in this paper.

Methods.—The x-ray radiation was applied by a Universal Type Coolidge tube with a tungsten target manufactured by the Victor X-Ray Corporation. A current of 8–9 milliamperes at 95 kilovolts was used to produce the radiation. The rays, filtered through a half-millimeter aluminum filter, had a wave-length of 0.72 Angström units.[2] The dosage was measured by a Fricke-Glaser dosimeter. At a distance of 18 centimeters from the target about 8 minutes were required to deliver 1000 r-units.

Young virgin females were x-rayed. The flies were placed into small vials (15 millimeters in diameter) filled with cotton and blotting paper up to five millimeters from the top edge of the vial. Etherized flies were placed in the vial which was then covered with one layer of cheese-cloth. Exposure was made through the cheese-cloth. The chamber of the dosimeter was covered with the same thickness of cheese-cloth as the vials to compensate for the absorption of rays by the cover of the vial.

Experiments.—To determine the relationship between the x-ray dosage and the frequency of primary non-disjunctions, two experiments were performed. They were numbered 14 and 19, respectively. In experiment 14, the constitution of the x-rayed female was *cv mt*-1 *sb*/*mt*-β and in the experiment 19 the females were of the *mt*-1 *f*-2/*mt*-β constitution. For each dosage tested, about 50 females were x-rayed in each one of the experiments. In experiment 19, for each dosage one-half of the females were x-rayed independently of the other half. In that way several independent x-rayings for certain dosages were obtained, giving evidence on the variability of the results.

The data of the experiments 14 and 19 are given in table 1, and they are graphically presented in figure 1. In figure 1 the result of each x-raying was shown separately; those of the experiment 14 are marked by a cross and those of experiment 19 by a circle. Average of all exposures for each dosage was marked by a full circle. Full line curve connects

TABLE 1

DATA SHOWING THE RELATIONSHIP BETWEEN THE INCREASE IN X-RAY DOSAGE AND THE NUMBER OF PRIMARY NON-DISJUNCTIONS

TREATMENTS	EXPERIMENT NUMBER	DATE OF TREATMENT	NUMBER OF CULTURES	REGULAR OFFSPRING ♀	♂	TOTAL	NON-DISJUNCTIONS ♀	♂	PER CENT OF MALE NON-DISJUNCTIONS
Control	14		118	9,607	6,190	15,797	–	3	0.0485
	19		48	3,973	3,946	7,919	1	11	.2788
Total			166	13,580	10,136	23,716	1	14	.1381
400 4-r-units	19	May 15	15	1,536	1,420	2,956	–	10	.704
	19	May 16	7	612	578	1,190	1	5	.865
Total			22	2,148	1,998	4,146	1	15	.751
500 r-units	14	March 31	47	5,182	4,271	9,453	1	31	.726
600 r-units	19	May 15	12	949	890	1,839	–	5	.562
	19	May 16	12	1,059	1,048	2,107	1	6	.573
Total			24	2,008	1,938	3,946	1	11	.568
800 r-units	19	May 15	12	988	909	1,897	1	8	.88
	19	May 16	6	504	447	951	–	3	.671
Total			18	1,492	1,356	2,848	1	11	.811
1000 r-units	14	March 29	23	2,427	2,304	4,731	1	23	.998
	14	March 31	26	2,152	1,968	4,120	1	28	1.423
	19	May 15	14	1,109	1,074	2,183	–	14	1.304
	19	May 16	10	895	863	1,758	3	9	1.043
Total			73	6,583	6,209	12,792	5	74	1.191
1200 r-units	19	May 15	13	1,140	1,044	2,184	–	12	1.149
	19	May 19	10	734	692	1,426	2	14	2.023
Total			23	1,874	1,736	3,610	2	26	1.498
2000 r-units	14	March 29	40	3,623	3,114	6,737	1	44	1.413
	19	May 15	12	885	800	1,685	1	7	.875
	19	May 19	7	483	428	911	1	9	2.103
Total			59	4,991	4,342	9,333	3	60	1.382
3300 r-units	14	March 21	41	1,537	1,262	2,799	1	22	1.74
4000 r-units	19	May 19	17	79	58	137	–	—	———
	19	May 22	42	313	266	579	1	5	1.880
	19	May 24	72	710	635	1,345	1	11	1.732
Total			131	1,102	959	2,061	2	16	1.668

averages. In table 2 data are given on the effect of x-ray treatment on the fertility of females. Those data were collected in connection with

TABLE 2

DATA SHOWING THE RELATIONSHIP BETWEEN X-RAY DOSAGE AND THE NUMBER OF OFFSPRING PER TREATED FEMALE

TREATMENT	PARENT FEMALES	NUMBER OF OFFSPRING	OFFSPRING PER ONE FEMALE
Control	65	7931	122
400 r-units	32	4162	130
600 r-units	29	3950	136
800 r-units	28	2860	102
1000 r-units	29	3967	137
1200 r-units	28	3638	130
2000 r-units	30	2614	87
3000 r-units	15	941	63
4000 r-units	135	2079	15.4

experiment 19. The broken line in figure 1 presents those data graphically.

Discussion.—As it may be seen from figure 1, the increase in the frequency of non-disjunctions for dosages up to 2100 r-units is rapid and is

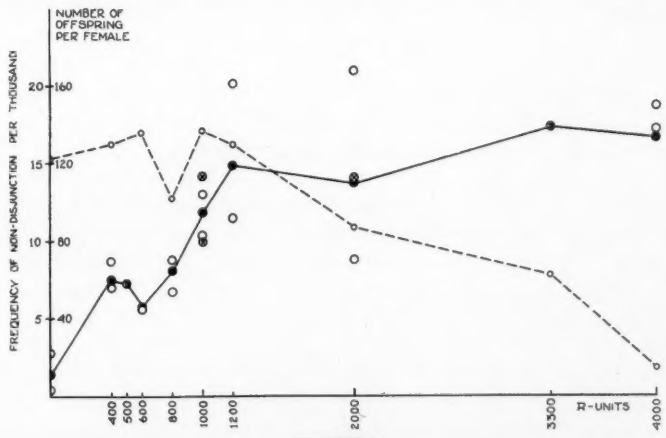


FIGURE 1

Relation between the x-ray dosage and the frequency of male non-disjunctions (solid line) and between the x-ray dosage and the number of offspring per treated female (broken line).

proportional to the increase in r-units. Between 1200 and 2000 r-units, however, the curve shows a sharp break indicating a decreased rate in the increase of non-disjunctions. It is of interest to note that at about the same point the number of offspring per treated female begins to decrease as indicated by the broken line curve. This coincidence suggests that the

decrease in the number of offspring per treated female may be responsible for the inflection in the non-disjunction curve, namely, that the offspring which would have shown non-disjunctions is killed off by x-rays in a larger proportion than the regular offspring.

Summary.—At low dosages the increase in the percentage of primary non-disjunctions was found to be almost proportional to the x-ray dosage applied. Between 1200 and 2000 r-units the curve shows a sharp break, the increase in the percentage of non-disjunctions becoming smaller. Since at about the same point the fertility of treated flies begins to decline, it is suggested that at higher dosages the offspring which would have shown non-disjunctions is killed off by x-rays in a larger proportion than the regular offspring.

[1] This investigation has been aided by a grant from the NATIONAL RESEARCH COUNCIL, Committee on the Effects of Radiation upon Living Organisms.

[2] Authors are indebted to Dr. H. Fricke for the calculation of the wave-length.

RECOVERY FOLLOWING GENETIC DEFICIENCY IN MAIZE[1]

BY L. J. STADLER

UNIVERSITY OF MISSOURI AND U. S. DEPARTMENT OF AGRICULTURE,
COLUMBIA, MISSOURI

Communicated October 11, 1930

Deficiency in Maize.—Chromosomal or sectional deficiency normally occurs with low but measurable frequency in early endosperm development in maize. This is shown in seeds heterozygous for linked endosperm characters by the occasional appearance of endosperm chimeras, in which a portion of the endosperm shows the loss of linked dominant characters present in the remainder. Of the 10 maize chromosomes, 7 may be marked by genes for endosperm characters, five chromosomes by one gene each (*R*, *Su*, *Y*, *Pr* and *A*), one by two linked genes (*Y* and *Bh*), and one by three linked genes (*C*, *Sh*, and *Wx*). Endosperm chimeras for each of the 7 chromosomes have been found in untreated material. The normal frequency varies from about 9 chimeras per 1000 seeds for *A* to less than 2 per 1000 seeds for *Su*. Evidence from chimeras of the *C-Sh-Wx* group indicates that the deficient region is often (possibly always) only part of the chromosome.

X-ray treatment shortly after fertilization not only increases greatly the frequency of deficiency in the young endosperm, as has been reported previously,[2] but induces deficiency in the young embryo as well. In cultures heterozygous for a plant character, irradiation at this stage causes

the production of a small proportion of deficient plants showing the recessive character. These are usually very defective in growth and partially sterile. If the treatment is delayed until some days after fertilization, plant chimeras with deficient sectors result.[3]

Irradiation of pollen also induces deficiency. In untreated *c sh wx* ears pollinated by x-rayed *C Sh Wx* pollen, the loss of the dominant genes in the pollen is shown by the production of a few seeds showing the recessive characters. Most of these are colorless, shrunken, and waxy (*c sh wx*), showing the loss of all three dominant genes, but a considerable proportion have lost *C* (and usually *Sh*) without losing *Wx*. Thus, as in the case of the endosperm chimeras, the deficiency is often, or perhaps always, sectional. Some of the losses of the dominant genes may be due to recessive gene mutation, but if so the proportion is very small. Similar loss of the other endosperm genes is shown in appropriate crosses. Although the distinction between deficiency and recessive mutation cannot be established individually for these cases, because of the lack of linked endosperm genes, it is assumed that these also are due largely to deficiency.

In a similar manner the loss in treated pollen of various dominant genes for plant characters is shown in the plants of the next generation. Deficiencies of 5 genes determining plant characters, *A*, *B*, *Lg*, *J* and *G*, have been found following x-ray treatment. The gene *A*, in the presence of certain complementary genes, determines both endosperm color and plant color. In the cross $a \times A$, treatment of mature pollen induces deficiency of *A* independently in endosperm and progeny, as expected from the fact that the two sperms which are to fuse respectively with the polar nuclei and the egg are separate in the mature pollen grain. When the treatment is applied several days earlier (before the division of the generative nucleus), the seeds with colored endosperm produce colored plants and those with colorless endosperm produce colorless plants.

Recovery in Deficient Endosperms.—Reversion following deficiency was first found in deficient endosperms resulting from pollen treatment. Tassels of an *A R C Wx Pr* stock were treated at various stages of development subsequent to microsporogenesis. As the pollen matured (from 1 to 16 days after treatment) pollinations were made on *a R C pr*[4] and *A R c wx* stocks to detect the loss of the dominant genes *A*, *Pr*, *C* and *Wx* in the pollen. Among 3916 seeds produced by pollination on *a R C pr* ears, 27 were colorless, indicating loss of *A*, and 9 were red, indicating loss *Pr*. In addition, 18 seeds were colorless except for one or more small spots of colored tissue, and 4 were red except for similar spots of purple. The portion of the seed showing the dominant character was most frequently a single spot considerably less than 1 sq. mm. in area. Under low magnification it is seen to consist of a group of aleurone cells ranging in number from about 20 to several hundred. Similar but

smaller spots hardly visible to the unaided eye appear on magnification. In many cases there are two or more spots on a single endosperm, and sometimes as many as 8 or 10. The frequency of endosperms with two or more spots is distinctly higher than would be expected from chance coincidence. On the average the total area showing the dominant character is less than $^1/_{100}$ of the area of the endosperm surface.

If the loss of *A* and *Pr* may be ascribed to recessive mutation, the reappearance of the dominant character may be considered the result of reverse mutation in a cell from which the dominant sector is derived. Since linked endosperm characters are not available in the *A* and *Pr* chromosomes, a positive distinction between deficiency and recessive mutation cannot be established for these cases. But results with the *c wx* crosses show that in these at least the loss and recovery are a chromosomal rather than a genic phenomenon. Among 2374 seeds produced by pollination of *A R c wx* ears with pollen of the treated plants, 27 were colorless (18 colorless waxy and 9 colorless non-waxy). In addition, 5 seeds were colorless waxy except for colored spots similar to those described above. In all 5 cases the colored spots were non-waxy, showing that the chromosome or section carrying *C* and *Wx*, which was absent or inactive in the remainder of the endosperm, was present and normally functional in the cells comprising the small dominant sector. Presumably each spot was derived from a cell in which the deficiency was somehow nullified. The return to normal activity of genes previously deficient is termed "recovery."

An endosperm largely deficient but with a small non-deficient sector might conceivably be a chimera of the type described in the first paragraph, resulting from the loss of a chromosome or section in an early division in endosperm development. Normally a loss occurring in the first division would produce a chimera showing the deficiency in only about half of the endosperm, and later losses would produce smaller deficient sectors. But it is possible that a first division loss may sometimes be followed by such asymmetric development that the resulting sectors of the mature endosperm are very different in size. Among endosperm chimeras in untreated material a few cases are found in which the dominant sector includes only about one-fourth, one-eighth or even a smaller fraction of the entire endosperm. Possibly these are the result of asymmetric development following a chromosomal irregularity in the first division. May not the seeds described above as examples of reverted deficiency be accounted for as more extreme cases of the same sort?

This explanation is inadequate for several reasons. (1) The frequent occurrence of the non-deficient tissue in several unconnected sectors is difficult to account for on the hypothesis of asymmetric development of a first division chimera. If the non-deficient sectors result from the re-

covery of the deficient genes after several cell generations, these may be regarded simply as cases in which recovery occurred in several cells of the same endosperm. (2) The frequency of the deficient seeds with small non-deficient sectors is far too high in proportion to the frequency of other endosperm chimeras on the same ears. The number of such seeds was in fact much greater than that of typical first division chimeras (seeds with approximately half of the endosperm showing deficiency), the respective frequencies being 27 and 12 among the progenies from the tassel treatments described. In untreated material, among more than 100,000 seeds examined (which included 35 typical first division chimeras for *A*, *Pr* or *C Wx*), only 6 deficient seeds with small non-deficient sectors have been found. (3) The frequency of typical endosperm chimeras (including all those in which the deficient sector occupies approximately one-half of the endosperm or less), is not materially higher in seeds produced by the use of x-rayed pollen than in untreated seeds. In other words, although pollen treatment results in a great increase in the frequency of deficient endosperms and endosperms deficient except for a small non-deficient sector, it does not materially affect the frequency of deficiencies presumably arising in the early divisions in endosperm development. There is, however, a moderate but distinct increase in the frequency of the anomalous chimeras in which the dominant sector occupies only one-fourth or one-eighth of the endosperm. This suggests that such chimeras may be, at least in part, the result of induced deficiency in the pollen followed by recovery in early endosperm development.

Recovery in Endosperm Chimeras.—Small areas of non-deficient tissue, similar to those resulting from recovery in deficient endosperms, are sometimes found also in the deficient sectors of typical endosperm chimeras. It is possible that a surface appearance suggesting recovery may sometimes result from the normally irregular process of endosperm development. The sectors in typical endosperm chimeras sometimes have very devious boundaries, and occasionally a narrow peninsula of one sector extends far into the other. A similar outgrowth passing through the interior of the endosperm could, in the case of a chimera for aleurone color, produce an island of colored tissue within the borders of a colorless sector. In the case of deficiencies involving *Wx*, it is possible (by the use of an iodine stain) to trace the outlines of the sectors within the interior of the endosperm as well as upon its surface. In several cases small spots of *C Wx* tissue have been found in *c wx* sectors, and in no instance was there any indication of an internal connection with the large non-deficient portion of the endosperm. In order to account for these spots as displaced portions of the unaffected tissue it would be necessary to postulate some kind of nuclear migration, often occurring at a rather late stage in endosperm development. In view of the evidence for recovery in deficient endo-

sperms, it is much more probable that the small areas of non-deficient tissue found in the deficient sectors of chimeras are the result of a similar recovery of deficient genes.

In the pollen treatment experiments summarized above, the 5 instances of recovery for the *C Wx* chromosome were found among 22 seeds in which both *C* and *Wx* were deficient, and none was found among 9 seeds deficient for *C* and not for *Wx*. This leaves open the possibility that recovery may be limited to cases in which the entire chromosome is deficient. But in a later experiment with pollen treatment, recovery was found in two endosperms deficient for *C-Sh* and not for *Wx;* and in the deficient sectors of endosperm chimeras, several similar instances have been found. In general, recovery appears to follow the deficiencies known to be sectional as frequently as the deficiencies which may involve the entire chromosome.

Recovery in the Sporophyte.—The cases thus far cited involve recovery only in the endosperm. Unless the phenomenon occurs also in the sporophyte it can of course have no effect in inheritance. None of the endosperm genes involved can be detected with certainty in the embryo, in the stocks used in these experiments, but the presence of *A* in the sporophyte can be determined by the anthocyan coloration of the plants of the next generation. In the absence of *A*, anthocyan is entirely lacking. In the cultures used in the study of deficiency and recovery of *A* following pollen treatment, the presence of the complementary plant color factor *B* insured the development of distinct red plant color on all exposed parts (except leaf blades) of the non-deficient plants.[5] From the seeds produced by the use of pollen treated at various stages, 1011 plants were grown, of which 9 were deficient for *A*, and one (plant 10–377.1–1) showed clear evidence of recovery. This plant was wholly green except for several narrow but distinct and sharply delimited stripes or lines of red on the lower leaf sheaths. The total area of the red portion was about $^1/_{50}$ of the area of the affected sheaths, and no red color was found above the fifth leaf. Like most deficient plants, plant 10–377.1–1 was very defective in growth, and in spite of optimum growing conditions it reached only about one-fourth the normal height, tasseled and silked late, and matured no pollen. Pollinated by an "a-tester" (*a C R*), it produced 8 wholly colorless seeds.

Recovery in Untreated Material.—Recovery is not limited to deficiencies induced by irradiation. Several thousand of seeds heterozygous for various endosperm characters (10,590 for *A* and *Pr*, 36,919 for *R*, and 92,551 for *C Wx*) were produced without irradiation in 1929, as material for study of the natural occurrence of endosperm chimeras. These were grown in isolated fields more than a mile distant from that in which the x-ray treatments were applied. Some 700 endosperm chimeras were

found, including for each of the chromosome regions several cases in which one or more small spots of non-deficient tissue occurred in a deficient sector. In addition, as perhaps more convincing evidence of deficiency and recovery, there were 11 seeds in which the entire endosperm was deficient except for one or more small spots of non-deficient tissue (1 seed for *A*, 5 for *R*, and 5 for *C Wx*). In proportion to the total frequency of deficient endosperms, the frequency of deficient endosperms showing recovery was as high in this untreated seed as in the seed produced by the use of x-rayed pollen.

Discussion.—Recovery, therefore, is not a consequence of x-ray treatment or of some special type of deficiency induced by x-rays; it is found more frequently in irradiated material merely because deficiency is more frequent in irradiated material. Recovery occurs in both "normal" and "induced" deficiencies, in deficiencies involving each of the four chromosome regions tested, in deficiencies which are certainly sectional as well as in those which may involve the entire chromosome, and in both sporophyte and endosperm tissue. These facts suggest that the germinal change which results in deficiency may commonly be an event somewhat less decisive than the absolute loss of a portion of the gene-chain. If deficiencies in general are reversible, a mechanism may be provided for the occurrence of apparent dominant "return" mutations without any actually "creative" change in the germ plasm. The possibility of an analogous basis for the behavior of the so-called mutable genes and eversporting characters also merits investigation.

The cytological mechanism resulting in deficiency and recovery is unknown. The term deficiency was applied originally by Bridges[6] to the loss or inactivation of a section of the chromosome, and the definition was later extended to include cases in which an entire chromosome is absent, "clearly due to loss and not to inactivation" (Morgan, Bridges and Sturtevant).[7] In the light of our present knowledge of chromosome behavior it seems extremely improbable that a chromosome may reappear in the progeny of a cell from which that chromosome is absent, but not inherently unlikely that an inactive chromosome may sometimes return to its active state. There is, however, no evidence at present for the existence and continued transmission of genetically inactive chromosomes. Moreover, Dr. Randolph's cytological study of maize plants in which deficiency had been induced by x-ray treatment of either the embryo or the parental pollen showed that these deficiencies commonly involve the absence of a chromosome or of a cytologically detectable chromosome fragment.[8]

A hypothetical mechanism, providing for recovery in deficient tissue in which cytological examination shows the absence of a chromosome, may be outlined as follows: Assume that the radiation or other stimulus

does not cause the destruction or loss of the chromosome, but merely deprives it of the power of reproduction. At each succeeding mitosis this chromosome goes to one daughter cell or the other. It is not necessarily inactivated as a functioning cell-constituent; but so long as it does not divide it will be present in only one cell of the organism, and its activity there will produce no visible effect. If in the course of development it recovers the power of reproduction, the progeny of the cell in which the recovery occurs will constitute a non-deficient sector. If recovery is gradual, the chromosome dividing sporadically for a time before regaining normal regularity, several non-deficient sectors may result. If recovery does not occur or is too long delayed, the individual will be typically deficient, and cytological examination of all but a negligible fraction of its cells will indicate that the chromosome has been lost. This hypothesis is subject to various experimental tests, which have not yet been completed.

Summary.—1. Recovery is the return to normal activity of genes previously deficient.

2. In endosperms, or in sectors of endosperm chimeras, in which deficiency has been induced by x-ray treatment, recovery is manifested by the occasional appearance of small spots of non-deficient tissue.

3. Deficient endosperms and endosperm chimeras occur also (though rarely) in untreated material. Recovery occurs as frequently in the deficient tissue of untreated material as in deficient tissue produced by irradiation.

4. Four chromosome regions (marked by the genes *A*, *R*, *C-Wx* and *Pr*) have been used extensively in experiments on deficiency in endosperm tissue. Several instances of recovery have been found for each of the four regions.

5. Recovery occurs in the sporophyte as well as in the endosperm. Deficiency of *A* in x-rayed pollen was followed in one case by recovery in the young sporophyte, producing a plant deficient except for several narrow stripes of non-deficient tissue.

[1] These experiments were supported in part by a grant from the National Research Council, Committee on Effects of Radiations on Living Organisms.

[2] Stadler, L. J., These PROCEEDINGS, **14**, 69–75 (1928).

[3] Stadler, L. J., *J. Hered.*, **31**, 3–19 (1920).

[4] For the complementary action of the genes *A*, *C*, *R*, and *Pr*, see Emerson, R. A., *Cornell Agr. Expt. Sta. Memoir*, **16** (1918).

[5] For the genetics of plant color in maize, see Emerson, R. A., *Ibid.*, **39** (1921).

[6] Bridges, C. B., *Genetics*, **2**, 445–465 (1917).

[7] Morgan, T. H., C. B. Bridges, and A. H. Sturtevant, *Bibliographia Genetica*, **2**, 3–262 (1925).

[8] Randolph, L. F., in a paper presented before Section O, *Amer. Assoc. Adv. Sci.*, Des Moines, 1929 (unpublished).

MAGNETIC ANALYSIS OF NEGATIVE IONS IN MERCURY VAPOR

BY WALTER M. NIELSEN*

BARTOL RESEARCH FOUNDATION, FRANKLIN INSTITUTE

Several years ago the writer[1] presented evidence which he regarded as indicating the formation of negative ions in mercury vapor. The method employed was essentially that previously used in a similar study by Mohler,[2] in which a beam of electrons was projected parallel to the axis of a tube and parallel to a magnetic field produced by a co-axial solenoid. Either positive or negative ions were swept out of the path of the beam by means of a cross electric field of appropriate sign applied between two electrodes symmetrically placed on two sides of the electron beam. The results were interpreted as indicating the attachment of slowly moving electrons to normal mercury atoms. At certain velocities of the electrons in the beam, some of which correspond to known critical potentials of the atom, the number of negative ions produced was observed to be much greater than would be expected from a consideration of the efficiency of excitation in mercury vapor[3,4] and the probability of electron attachment as indicated by the data obtained at small electron speeds. This fact made it appear as though electron attachment was greater to an excited atom than to a normal atom.

The writer has recently undertaken a study of the character of such negative ions by a magnetic analysis of the ions produced by a discharge in mercury vapor and the present report is concerned with that study.

Method of Measurement.—The experimental arrangement sketched in figure 1 is, in some respects, similar to that employed by others in this field.[5]

The essential change was a rotation of the filament assembly so that the electrons from a tungsten filament F are projected parallel to a magnetic field of approximately 100 gauss, produced by a pair of Helmholtz coils as indicated in figure 1*b*. The ions were produced by a beam of electrons indicated by the dots between A and B of figure 1*a*. The average electrical potential of the region occupied by the beam could be adjusted equal to that of the accelerating grids by obvious methods. The total electron current was measured by a galvanometer connected to the cylinder G. Secondary electrons were in part prevented from entering the beam by applying a potential of approximately 100 volts between G and an insulated electrode inside of G.

The filament leads and accelerating grids were mounted on a removable press. Picein wax on the ground glass joint, which held this press, was cooled with tap water by means of a coil wrapped around it. The cylinder

G and electrode A were supported by stiff molybdenum wire, spot-welded to tungsten-pyrex seals through an intermediary piece of tantalum.

Mercury vapor from a well (not shown) was partially collimated by a narrow slit (1 mm. $\times$ 10 mm. $\times$ 25 mm.). It passed up through the discharge space, and condensed on the walls of a tube cooled by carbon dioxide snow or liquid air.

Positive or negative ions formed in the beam were pulled out by means of a potential (V_c) between A and B. Some of the ions passed through a slit (2 mm. $\times$ 25 mm.) in the plate covering the end of cylinder B. Between B and the cylinder C, the ions were again accelerated (V_a), and some of them passed through the two slits (1 mm. $\times$ 20 mm.) in the ends of the otherwise closed cylinder C. A transverse magnetic field was applied in the region D by placing the curved portion of the tube between the poles of an electromagnet as indicated by the dotted lines. Ion currents due to ions which passed through a final slit in J and fell on a plate at the end of the tube were measured with an electrometer by the rate of deflection method. For the negative ions a sensitivity of 1200 mm./volt was employed; for the positive ion the sensitivity was reduced to a convenient value by altering the potential of the needle.

All metal parts were molybdenum except the metal shield near the wall of the analyzing chamber and the slitted rod attached to H, which were of aluminum. A baking out at 450°C. was given all metal parts before assembly. During a run the pressure of non-condensable gases was not detectable on a McLeod gauge reading to 10^{-5} mm. of mercury.

It became evident in the preliminary experiments that a positive space charge in the beam was a cause of considerable difficulty. This was made apparent by the increased velocity of the positive ions as determined by the magnetic analysis of the ions. A retarding potential curve showing the number of ions reaching C as a function of a reverse potential between B and C also indicated a space charge in some cases as high as eight volts. Under such circumstances it was quite difficult to extract negative ions from the beam, apparently because the applied potentials were incapable of neutralizing the field gradient set up in the vicinity of the beam. Such difficulties were very much reduced at lower vapor pressure and at a pressure of 10^{-4} mm. of mercury, as finally used, the effects of space charge were quite negligible as indicated by the consistency of the data obtained in runs, in which different accelerating voltages were employed.

In the results presented, the temperature of the well was approximately 50° C. The pressure of the mercury vapor was estimated to be of the order of 10^{-4} mm. of mercury, from a knowledge of the total electron current, the positive ion current to B, the efficiency of ionization[6] and the length of electron path.

The procedure was to record the ion current for different values of Va

with a constant value of the magnetic field intensity, or as carried out in the curves displayed in the next section, to measure the ion currents as a functions of the magnetic field intensity with constant accelerating potentials.

Experimental Results and Discussion.—The results obtained are displayed in the curves of figure 2, in which the ratio of ion to total electron

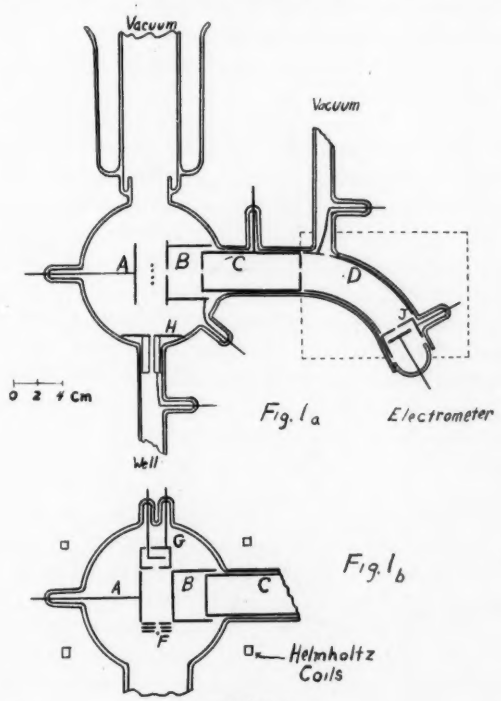


FIGURE 1
Diagram of apparatus.

current to G is plotted along the ordinate and the ratio of the mass (m) of the charged particle to the charge (e) carried by the particle is plotted along the abscissa. Such effective values of m/e were calculated from a knowledge of the radius of curvature of the ion path, the applied voltages and the magnetic field intensities as measured by a Grassot fluxmeter. In these runs, the drawing out potential (Vc) was 20.2 volts and the accelerating potential (Va) for ions was 54.8 volts. The driving potential for electrons was 70 volts. The maximum value of the magnetic field intensity employed in the runs was 2000 gauss. The height of the Hg^+

peak has been reduced by dividing the ordinate of the three points near the maximum by a factor of two so as to make evident some of the smaller peaks.

The negative ion currents observed by reversing all accelerating voltages and the direction of the magnetic field are indicated by the crosses. The magnitude of the negative ion currents should be divided by a factor of 800 in comparing them with the positive ion currents.

We ascribe to mercury the negative ion peak which appears at a value of m/e equal to 200. Within the accuracy of the determination it has the same effective value of m/e as Hg^+. The pressure range over which the apparatus was capable of giving reliable data was limited at higher pressures by a positive space charge as already indicated. It is also of interest to point out that it was necessary to cool the jacket for several hours before beginning a run in order that the space between the slits and the analyzing region was relatively free of mercury vapor. In this respect the behavior of the negative ions was markedly different from that of the positive ions. Apparently such ions are either of low stability or have a large collision area or possibly both.[7]

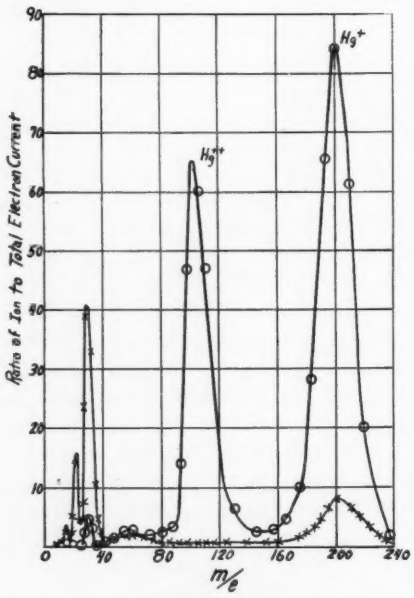


FIGURE 2

Analysis of positive and negative ions.

The ion which has a value of m/e equal to 29 is probably due to carbon monoxide ($m/e = 28$) since the resolving power of the apparatus was not sufficient to separate ions for which m/e differs by one unit in this region. This impurity is quite frequently observed in work on positive ion analysis.[6,8] The ion having an m/e value of 12 is undoubtedly due to carbon, probably produced by ionization of carbon monoxide. A similar ion is observed in the work of Hogness and Harkness,[9] in their work on the ionization of carbon monoxide.

The ions which have effective values of m/e of 19 and 60 have not been identified with certainty. Water vapor is suggested as the origin of the former.

All of the above ions appear in the analysis of the positive ions as well

as in the negative ion analysis. The failure of others who have made a search for negative ions, as here reported, may have been due to the formation of positive ion sheaths.[5] It is very possible that the presence of such sheaths may throw doubt upon the interpretation of results obtained by positive ion analysis, unless such sheaths are recognized or avoided. It is also necessary in considering the ability of the present apparatus to detect negative ions to recall that slow electrons present in the beam as a result of inelastic collisions or as secondary electrons are constrained to remain there by the collimating magnetic field for a larger period of time than in the methods used heretofore. That slowly moving electrons are present in the electron beam is indicated by the abundance of the CO^- and CO^+ ions as determined by the areas under their respective peaks. An inspection of the curves leads to the conclusion that the number of CO^- ions is approximately 1/100 of the number of CO^+ ions, a ratio much greater than would be expected from a consideration of the magnitude of the attachment constant (average number of impacts between electrons and molecules of a gas which result in one attachment to form a negative ion) and the efficiency of ionization of similar gases, unless a large number of slowly moving electrons are present in the beam. Because of this fact, a calculation of the attachment constant for mercury from the known pressure total electron current and negative ion current becomes meaningless. We may, however, avoid this difficulty, and, by making what appear to be fairly justifiable assumptions, calculate a value for the attachment constant for mercury.

We first assume that the efficiency of ionization of carbon monoxide to be of the same order of magnitude as that of N_2[10] at the same electron energy. A knowledge of the abundance ratio of Hg^+ and CO^+ ions may be obtained by a measurement of the respective peak areas properly corrected to take account of the effective slit width for the two types of ions.[6,11] This abundance ratio, together with the known partial pressure of mercury, the efficiency of ionization of mercury, and the assumed efficiency of ionization of carbon monoxide lead, by a simple calculation to a value of 3.6×10^{-6} mm. of mercury for the partial pressure of CO present as an impurity in the experimental tube.[12]

Our knowledge of the attachment constants are limited to pressures very much greater than those employed here, and because of lack of knowledge of the mechanism of the attachment process, the effect of de-attachment, etc., an extrapolation to a very low pressure may not be justified. It is of interest, however, to make such an extrapolation and from the ratio of ion currents, the partial pressures of mercury and carbon monoxide, as estimated above, to calculate the order of magnitude of the attachment constant for mercury. The value[13] of the attachment constant n for carbon monoxide is 1.6×10^8.

The above value of n is given as being accurate in order of magnitude only. If one assumes the abundance ratio of the Hg^- and CO^- ions to be inversely proportional to the attachment constant, a simple calculation leads to a value of 6×10^{10} for the attachment constant of mercury.[14] Obviously no great confidence can be placed in the absolute value of this attachment constant. It is presented here as a qualitative estimate and as such may indicate the order of magnitude in a comparison with other gases.

In conclusion the writer would like to express his indebtedness to Dr. Swann and the Board of the Bartol Foundation for the privilege of working at the Laboratory of the Foundation, and to the National Research Council for the opportunity to carry on this work.

* NATIONAL RESEARCH FELLOW.

[1] Walter M. Nielsen, *Phys. Rev.*, **27**, 716 (1926).

[2] F. L. Mohler, *Ibid.*, **26**, 614 (1925).

[3] H. Sponer, *Zeits. f. Physik.*, **7**, 185 (1921).

[4] G. Hertz, *Ibid.*, **32**, 298 (1925).

[5] See R. W. Gurney and P. M. Morse, *Phys. Rev.*, **33**, 789 (1929).

[6] Walker Bleakney, *Ibid.*, **35**, 139 (1930).

[7] See H. Bethe, *Zeits. f. Physik*, **57**, 815 (1929), for a wave mechanical calculation of the charge distribution for negative atomic hydrogen ions.

[8] H. Barton and J. H. Bartlett, *Phys. Rev.*, **31**, 822 (1928).

[9] T. R. Hogness and R. W. Harkness, *Ibid.*, **32**, 936 (1928).

[10] K. T. Compton and C. C. Van Voorhis, *Ibid.*, **27**, 724 (1926). While this assumption is admittedly an approximation, the qualitative conclusion which follows later as to the order of magnitude of the attachment constant for mercury very probably will not be affected by this uncertainty.

[11] Let ρ be the radius of curvature of the ionic path as defined by the slit system and let $d\rho$ represent the width of the slit at J. An ordinate of figure 2 corresponding to a particular value of m/e (or to a particular value of H^2 since $\frac{\rho^2 H^2}{2Vc^2} = \frac{m}{e}$, where the symbols have their usual meaning) is a measure of the number of ions having radii between ρ and $\rho + d\rho$. The area under any peak is then equal to the value of the integral

$$A_1 = \int_{H_1^2}^{H_2^2} \frac{dn}{d\rho}\, d(H^2),$$

where n is the total number of charges with paths having radii between ρ_1 and ρ. The integral required in order to compare the abundancies of two different types of ions is

$$I_1 = \int_{H_1^2}^{H_2^2} \frac{dn}{d\rho} \frac{d\rho}{d(H^2)}\, d(H^2).$$

If we assume that $\frac{d\rho}{d(H^2)}$ is a constant over the region of integration corresponding to a single peak

$$I_1 = \frac{d\rho}{d(H^2)} A_1.$$

From the equations of motion we have

$$H^2\rho^2 = \frac{2c^2 V m_1}{e_1} \text{ and } \frac{d\rho}{d(H^2)} = \frac{\rho}{H^2} = (\text{constant}) \frac{e_1}{m_1},$$

and therefore

$$I_1 = A_1 (\text{constant}) \frac{e_1}{m_1}.$$

We conclude that the areas of the peaks due to the different kinds of ions should be divided by the corresponding value of m/e in comparing ion abundancies.

[12] $\frac{I^+_{Hg}}{I^+_{CO}} = \frac{p_{Hg} P_{Hg}}{p_{CO} P_{CO}}$ where $\frac{I^+_{Hg}}{I^+_{CO}}$ represents the ratio of the abundance of positive mercury ions to positive carbon monoxide ions as measured, p_{Hg} and p_{CO} represent the partial pressures of mercury and carbon monoxide, respectively, and P_{Hg} and P_{CO} represent the probabilities of ionization by impact between electrons and mercury atoms and between electrons and carbon monoxide molecules, respectively.

[13] L. Loeb, *Kinetic Theory of Gases*, p. 513.

[14] $\frac{I^-_{CO}}{I^-_{Hg}} = \frac{n_{Hg} p_{CO}}{n_{CO} p_{Hg}}$ where $\frac{I^-_{CO}}{I^-_{Hg}}$ represents the ratio of the abundance of carbon monoxide negative ions to mercury negative ions, n_{Hg} and n_{CO} represent the attachment constants for mercury and carbon monoxide, respectively, and where p_{Hg} and p_{CO} are defined as in note 13.

THE PRESSURE OF THE WIND ON LARGE CHIMNEYS

BY HUGH L. DRYDEN AND GEORGE C. HILL

NATIONAL BUREAU OF STANDARDS, WASHINGTON, D. C.

Read before the Academy, September 18, 1930

Although the circular cylinder is one of the simplest geometrical forms, the phenomena attendant upon the flow of air around it are extremely complicated. It has not as yet been possible to give a complete theory of the motion of the air, but many properties of the flow have been determined by experiment for the range of sizes and speeds available in wind tunnels. The experiments which are here outlined were made in connection with the practical problem of determining the wind pressure exerted on a large chimney in high winds, and give additional information as to the flow around large cylinders.

For a certain range of sizes and speeds, it is found that the total force exerted by the wind is proportional to the area of the projection of the cylinder on a plane normal to the wind and to the velocity pressure (i.e., one-half the product of the density of the air by the square of the wind

speed). The ratio of the force to the product of projected area and velocity pressure is called the force coefficient. A more exact analysis shows that the force coefficient is a function of the so-called Reynolds Number, i.e., product of velocity by the diameter of the cylinder divided by the kinematic viscosity of the air. At the highest Reynolds Numbers reached in ordinary wind tunnels, the force coefficient is found to decrease very rapidly with increasing Reynolds Number, reaching a minimum whose value is one-third of the value at smaller Reynolds Numbers. In other words, the wind force on a large cylinder at a given speed is very much less than that which would be obtained by increasing the force on a small cylinder at the same speed in the ratio of the exposed areas, or alternatively, the force on a given cylinder at high speed is much less than that which would be obtained by increasing the force at low speed in the ratio of the squares of the speeds.

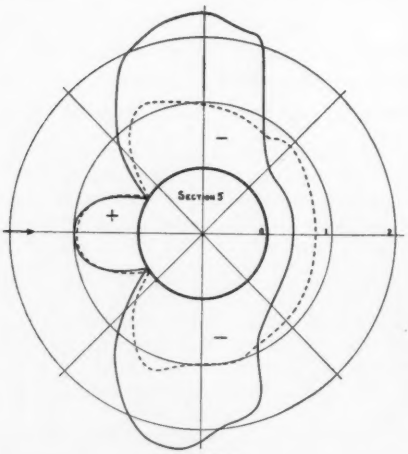


FIGURE 1

Distribution of pressure on 8-inch cylinder. Radial distances from the base circle are proportional to the ratio to the velocity pressure of the difference in pressure between the station and the atmospheric pressure. In the region marked +, the pressure on the cylinder exceeds the atmospheric pressure, while in the region marked —, the pressure on the cylinder is less than the atmospheric pressure. The wind direction is shown by the arrow. The broken line is for a speed of 40 ft./sec., the solid line for 80 ft./sec.

We have measured the changes in the distribution of the wind pressure around the cylinder occurring at this so-called critical region and some of the results are shown in figure 1. The heavy black circle represents the cylinder, with air approaching in the direction shown by the arrow. The changes in pressure from the normal atmospheric pressure which are produced by the wind are plotted radially outward. The region marked plus is a region of increased pressure, the region marked minus a region of decreased pressure. The unit for measuring the pressure is the velocity pressure. The dotted curve refers to a low speed (40 ft./sec.) and the solid curve to a high speed (80 ft./sec.) for a cylinder 8 inches in diameter. The extensive area under reduced pressure is surprising. A little study shows that in spite of the larger reductions in pressure at the sides, the solid curve corresponds to a lower force in the direction of the wind.

Attention is called to the relatively small change in the angle at which the pressure equals the normal atmospheric pressure as compared with changes in the magnitude of the reduction in pressure at the sides or at the rear.

To obtain information on cylinders comparable in size to large chimneys, it is necessary to supplement experiments in wind tunnels by experiments on large cylinders in natural winds. The principal difficulty in such experiments is the determination of the velocity of the wind. It occurred to one of us that it might be possible to secure the wind speed from the

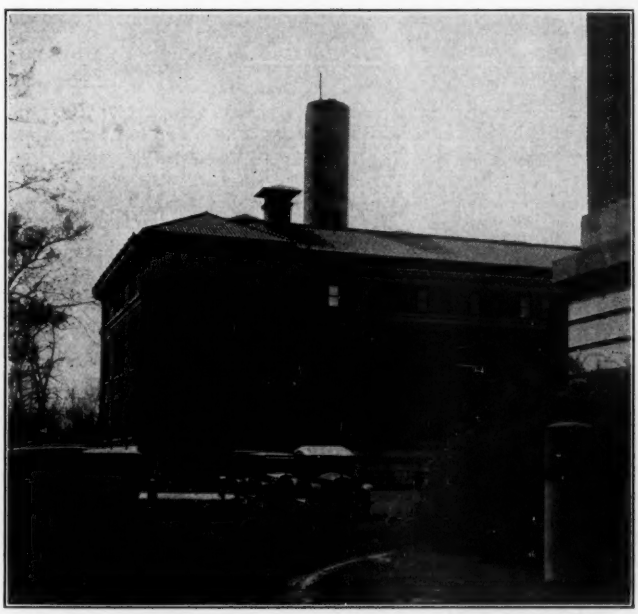


FIGURE 2

Experimental stack, 10 ft. in diameter, 30 ft. high, erected for measurements of wind pressure in natural winds.

pressure distribution diagram by estimating the angle at which the pressure is equal to the normal atmospheric pressure, and determining the maximum increase in pressure above the normal atmospheric pressure, thus obtaining the velocity pressure. It has already been pointed out that this angle is not as variable as other properties of the distribution.

Since the Bureau of Standards was to erect a new power plant chimney on which a more or less permanent installation could be made, it was considered advisable to try the new method on a somewhat smaller scale and

under such conditions that the natural wind measurements could be made to stand alone. For this purpose an experimental stack 10 feet in diameter and 30 feet high (as shown in figure 2) was erected on the roof of the West building of the Bureau. Pressures were measured at twenty-four stations around the circumference at a single elevation, about two-thirds of the height from the base of the stack. The wind speed was determined by means of a pitot-static tube mounted on a weather vane about ten feet higher than the top of the stack. The overturning moment was measured by mounting the stack on pressure capsules connected to pressure gauges. The value of the average wind pressure determined from the over-

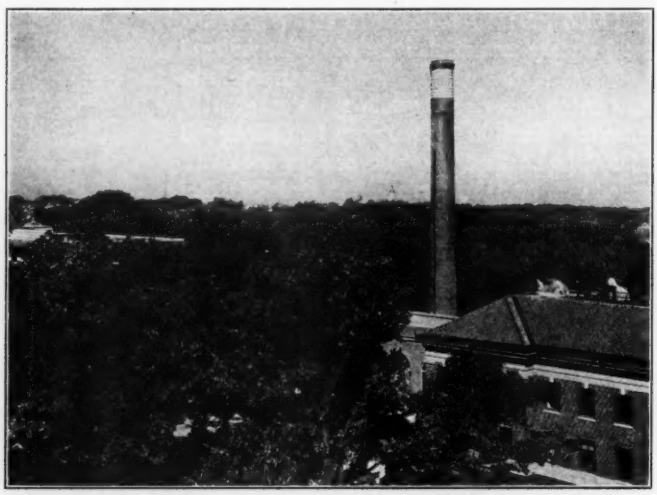


FIGURE 3

The power plant chimney of the Bureau of Standards.

turning moment was somewhat higher than the value determined from the pressures at a single elevation. The method of determining the velocity from the distribution itself was found to be reasonably satisfactory when the various attendant circumstances were considered, especially the small value of the length-diameter ratio of the experimental stack.

On the occasion of the construction of a new power plant at the Bureau of Standards, provision was made for measurements of the distribution of wind pressure on the stack at a single elevation so that some information might be obtained as to the wind pressure on an actual chimney. A view of the stack and its surroundings looking toward the northwest (the direction of the prevailing wind) is shown in figure 3. The dark

line running down the stack is the group of 24 pipes leading to the pressure gauge and the holes are located about 2 feet above the top of the pipes. The stack at this point is 11.8 feet in diameter.

Space does not permit any detailed statement or comparison of the results. A full account of the work is given in *Research Paper* **221** appearing in the September, 1930, issue of the Bureau of Standards *Journal of Research*. The general conclusions drawn from the tests may be stated as follows:

1. The wind pressure on a chimney at a given wind speed is a function of the ratio of the height of the chimney to its diameter and possibly also of the roughness of its surface.

2. Experiments on small cylinders cannot be directly used to predict the wind pressure on a full scale chimney because of the large scale-effect.

3. A wind pressure corresponding to 20 lbs. per square foot of projected area at a wind speed of 100 miles per hour is a safe value to use in designing chimneys of which the exposed height does not exceed 10 times the diameter.

4. The pressure may reach large values locally and this may need consideration in the design of thin-walled stacks of large diameter.

5. Further experiments are necessary to obtain satisfactory information as to the variation of wind pressure with the ratio of height to diameter.

PROJECTIVE NORMAL COÖRDINATES

BY L. P. EISENHART

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated October 3, 1930

A projective geometry of paths was formulated first by T. Y. Thomas,[1] and included the definition of a type of projective normal coördinates. Veblen and J. M. Thomas[2] proposed a system of such coördinates as solutions of a certain system of partial differential equations in n dependent variables. In §41 of my Non-Riemannian Geometry[3] I showed that the determination of these coördinates could be reduced to the solution of a differential equation in one dependent variable. It is the purpose of this note to give this equation another form and thence obtain the explicit form of the expressions of general coördinates in terms of projective normal coördinates.

1. Consider an n-dimensional space V_n of coördinates x^i. We take as the basic elements of V_n the *paths*, that is, the integral curves of the system of differential equations.

$$\frac{dx^i}{dt}\left(\frac{d^2x^j}{dt^2} + \Gamma^j_{kl}\frac{dx^k}{dt}\frac{dx^l}{dt}\right) - \frac{dx^j}{dt}\left(\frac{d^2x^i}{dt^2} + \Gamma^i_{kl}\frac{dx^k}{dt}\frac{dx^l}{dt}\right) = 0, \quad (1.1)$$

where Γ^i_{kl} are functions of the x's, and without loss of generality we take $\Gamma^i_{kl} = \Gamma^i_{lk}$. Throughout the paper we use the convention that when the same index appears as a subscript and superscript in a term, as in the case of k above, this term stands for the sum of terms as the index takes the values 1 to n. The independent variable t in (1.1) is a general one and equations (1.1) retain this form for any independent variable.

In order that a second set of functions $\bar{\Gamma}^i_{jk}$ define the same paths, we have on subtracting (1.1) from similar equations in the $\bar{\Gamma}$'s,

$$A^{ij}_{hkl}\frac{dx^h}{dt}\frac{dx^k}{dt}\frac{dx^l}{dt} = 0, \; A^{ij}_{hkl} = \delta^i_h\,(\bar{\Gamma}^j_{kl} - \Gamma^j_{kl}) - \delta^j_h\,(\bar{\Gamma}^i_{kl} - \Gamma^i_{kl}).$$

Since these equations must be satisfied for every path, we must have

$$A^{ij}_{hkl} + A^{ij}_{klh} + A^{ij}_{lhk} = 0. \quad (1.2)$$

Contracting these equations for i and h, we have

$$\Gamma^i_{kl} - \frac{1}{n+1}(\delta^i_k\Gamma^h_{hl} + \delta^i_l\,\Gamma^h_{hk}) = \bar{\Gamma}^i_{kl} - \frac{1}{n+1}(\delta^i_k\bar{\Gamma}^h_{hl} + \delta^i_l\bar{\Gamma}^h_{hk}). \quad (1.3)$$

When these conditions are satisfied, so also are equations (1.2). If we define functions π^i_{kl} by

$$\pi^i_{kl} = \Gamma^i_{kl} - \frac{1}{n+1}(\delta^i_k\Gamma^h_{hl} + \delta^i_l\Gamma^h_{hk}), \quad (1.4)$$

then the equations of the paths may be written

$$\frac{dx^i}{dt}\left(\frac{d^2x^j}{dt^2} + \pi^j_{kl}\frac{dx^k}{dt}\frac{dx^l}{dt}\right) - \frac{dx^j}{dt}\left(\frac{d^2x^i}{dt^2} + \pi^i_{kl}\frac{dx^k}{dt}\frac{dx^l}{dt}\right) = 0. \quad (1.1')$$

From (1.4) we have

$$\pi^i_{il} = 0, \quad (1.5)$$

and from (1.3) we obtain $\pi^i_{jk} = \bar{\pi}^i_{jk}$. The functions π^i_{jk}, which accordingly are independent of the choice of the Γ's subject to the relations (1.3), we call the *projective coefficients* of connection, following T. Y. Thomas.[1]

When the coördinates x^i are subjected to a general analytic transformation into coördinates x'^i and the equations of the paths in the new coördinates are written

$$\frac{dx'^i}{dt}\left(\frac{d^2x'^j}{dt^2} + \Gamma'^j_{kl}\frac{dx'^k}{dt}\frac{dx'^l}{dt}\right) - \frac{dx'^j}{dt}\left(\frac{d^2x'^i}{dt^2} + \Gamma'^i_{kl}\frac{dx'^k}{dt}\frac{dx'^l}{dt}\right) = 0,$$

it can be shown[2] that the Γ's and Γ''s are related by the equations

$$\frac{\partial^2 x'^\alpha}{\partial x^i \partial x^j} = \pi_{ij}^k \frac{\partial x'^\alpha}{\partial x^k} - \pi'^\alpha_{\beta\gamma} \frac{\partial x'^\beta}{\partial x^i} \frac{\partial x'^\gamma}{\partial x^j} + \frac{\partial x'^\alpha}{\partial x^i} \frac{\partial \theta}{\partial x^j} + \frac{\partial x'^\alpha}{\partial x^j} \frac{\partial \theta}{\partial x^i}, \quad (1.6)$$

where $\pi'^\alpha_{\beta\gamma}$ are defined in terms of $\Gamma'^\alpha_{\beta\gamma}$ by equations of the form (1.4), and where we have put

$$\theta = \frac{1}{n+1} \log \Delta, \quad \Delta = \left| \frac{\partial x'}{\partial x} \right|, \quad (1.7)$$

that is, Δ is the jacobian of the transformation.

For a coördinate system x^i a parameter x can be chosen for each path so that the equations of the path are

$$\frac{d^2 x^i}{dx^2} + \pi_{jk}^i \frac{dx^j}{dx} \frac{dx^k}{dx} = 0. \quad (1.8)$$

We call x the *projective parameter* of the path for the x's, following T. Y. Thomas.[1] By means of (1.6) we show that the projective parameter x' for the same path in the coördinate system x'^i is given by

$$x' = \int_0^x e^{2\theta} dx. \quad (1.9)$$

We consider all the paths through a point P of coördinates x_0^i and without loss of generality we choose the parameter x so that $x = 0$ at P. If we assign coördinates y^i at points of each path through P by means of the equations

$$y^i = \left(\frac{dx^i}{dx} \right)_0 x, \quad (1.10)$$

we find,[4] on expanding the solution of (1.18) as power series, that

$$x^i = x_0^i + y^i - \frac{1}{2} (\pi_{jk}^i)_0 y^j y^k \ldots - \frac{1}{r} (\pi_{j_1 \ldots j_r}^i)_0 y^{j_1} \ldots y^{j_r} + \ldots, \quad (1.11)$$

where the functions $\pi_{j_2 \ldots j_r}^i$ are determinate functions of π_{jk}^i and their derivatives of order $r - 2$ at most, and where $(\ )_0$ indicates the value at P.

Thus equations (1.11) are the equations of transformation of coördinates x^i and y^i. These equations obtain within a domain about P such that no two paths through P, which lie entirely within the domain, meet again in the domain.

The transformation (1.11) belongs to the class of transformations of coordinates related to the x's by equations of the form

$$x^i = x_0^i + y^i - \frac{1}{2} (\pi_{jk}^i)_0 y^j y^k + \varphi^i, \quad (1.12)$$

where φ^i are functions of the y's which with their first and second derivatives vanish when the y's are zero. We have called any set of y's so

defined *projective coördinates*,[5] and have shown that because of (1.5) we have

$$\left(\frac{\partial}{\partial y^i} \log \Delta\right)_0 = 0, \tag{1.13}$$

where Δ is the jacobian $\left|\frac{\partial x}{\partial y}\right|$.

From (1.10) we have

$$\frac{d^2y^i}{dx^2} = 0, \frac{dy^i}{dx} = \frac{y^i}{x}. \tag{1.14}$$

If we denote by $\bar{\pi}^i_{jk}$ the projective coefficients in the y's and remark that x is not the projective parameter for the y's, it follows from equations of the form (1.1′) in the y's and from (1.14) that

$$\frac{\bar{\pi}^i_{jk}y^jy^k}{y^i} = \frac{\bar{\pi}^l_{jk}y^jy^k}{y^l}. \tag{1.15}$$

Expressing the x's as functions of the y's, we have that equations (1.8) reduce to

$$\left(\frac{\partial^2x^i}{\partial y^j\partial y^k} + \pi^i_{hl}\frac{\partial x^h}{\partial y^j}\frac{\partial x^j}{\partial y^k}\right)y^jy^k = 0,$$

which because of equations of the form (1.6) are reducible to

$$\left(\bar{\pi}^h_{jk}y^jy^k + 2\,y^h\frac{\partial\theta}{\partial y^j}y^j\right)\frac{\partial x^i}{\partial y^h} = 0,$$

where θ is given by (1.7) with $\Delta = \left|\frac{\partial x}{\partial y}\right|$. Since Δ is not equal to zero, these equations are equivalent to

$$\bar{\pi}^h_{jk}y^jy^k = -2y^h\frac{\partial\theta}{\partial y^j}y^j. \tag{1.16}$$

2. If φ is any function of the y's and we put

$$z^i = \frac{y^i}{\varphi}, \quad \psi = \varphi - \frac{\partial\varphi}{\partial y^i}y^i, \tag{2.1}$$

we have

$$\frac{\partial z^i}{\partial y^j}y^j = \frac{y^i}{\varphi^2}\psi, \quad \frac{\partial\psi}{\partial y^j}y^j = -\frac{\partial^2\varphi}{\partial y^i\partial y^j}y^iy^j, \tag{2.2}$$

$$\frac{\partial^2z^i}{\partial y^j\partial y^k}y^jy^k = \frac{y^i}{\varphi^3}\left(\varphi\frac{\partial\psi}{\partial y^j}y^j - 2\psi\frac{\partial\varphi}{\partial y^k}y^k\right). \tag{2.3}$$

If $\bar{\Delta}$ denotes the jacobian $\left|\frac{\partial z}{\partial y}\right|$, we find that[6]

$$\bar{\Delta} = \frac{\psi}{\varphi^{n+1}}, \quad \bar{\theta} = \frac{1}{n+1} \log \frac{\psi}{\varphi^{n+1}}. \tag{2.4}$$

Denoting by P^i_{jk} the projective coefficients in the z's, we have from (2.3), from equations of the form (1.6) in the z's and y's, and from (1.16)

$$y^i y^j \frac{\partial}{\partial y^j} \log (\psi^{n-1}\Delta^2) + (n+1) P^i_{jk} z^j z^k . \psi = 0,$$

where $\Delta = \left|\frac{\partial x}{\partial y}\right|$. Hence, if we choose the function φ so that

$$\varphi - \frac{\partial \varphi}{\partial y^j} y^j = \Delta^{\frac{2}{1-n}}, \tag{2.5}$$

we have for the coördinates z^i the equations

$$P^i_{jk} z^j z^k = 0. \tag{2.6}$$

Because of (1.5) it follows from (1.11) that Δ is of the form

$$\Delta = 1 + a_{ij} y^i y^j + \ldots$$

Consequently, $\Delta^{\frac{2}{1-n}}$ is of the form

$$\Delta^{\frac{2}{1-n}} = 1 + b_{ij} y^i y^j + \ldots + b_{i_1 \ldots i_r} y^{i_1}_1 \ldots y^{i_r} + \ldots,$$

and this expansion holds for the domain about P within which $\Delta \neq 0$. A solution of (2.5) is given by

$$\bar{\varphi} = 1 - \left(b_{ij} y^i y^j + \ldots + \frac{1}{r-1} b_{i_1 \ldots i_r} y^{i_1} \ldots y^{i_r} + \ldots \right). \tag{2.7}$$

Making use of (1.10) we see that along each path through P the domain of convergence of the right-hand member of (2.7) is the same as for the above expression for $\Delta^{\frac{2}{1-n}}$. Consequently these two expansions have the same domain.

From the form of equation (2.5) it follows that the general solution of this equation is

$$\varphi = \bar{\varphi} + \tau, \tag{2.8}$$

where τ is an arbitrary homogeneous function of the first degree in the y's. At P we have $(\varphi)_0 = 1$ whatever be τ. Then from (1.11) and (2.1) we have

$$\left(\frac{\partial x^i}{\partial y^j}\right)_0 = \left(\frac{\partial z^i}{\partial y^j}\right)_0 = \left(\frac{\partial y^i}{\partial z^j}\right)_0 = \left(\frac{\partial x^i}{\partial z^j}\right)_0 = \delta^i_j. \tag{2.9}$$

If any solution φ of (2.5) is used in (2.1) to define coördinates z^i and we define a parameter z for each path by

$$z = \frac{x}{\varphi}, \tag{2.10}$$

we have, in consequence of (1.14) and (2.2),

$$\frac{dz}{dx} = \frac{\psi}{\varphi^2}, \quad \frac{dz^i}{dz} = \frac{dy^i}{dx}, \tag{2.11}$$

and because of (1.14)

$$\frac{d^2z^i}{dz^2} = 0. \tag{2.12}$$

In view of this result and (2.6) we have that z defined by (2.10) is the projective parameter for the coördinates z^i. Since $(\varphi)_0 = 1$ and $(\psi)_0 = 1$, as follows from (2.5), we have from (2.11) that $\left(\frac{dz}{dx}\right)_0 = 1$. Consequently we obtain from (1.10), (2.1) and (2.10)

$$z^i = \left(\frac{dx^i}{dz}\right)_0 z. \tag{2.13}$$

If for any solution φ of (2.5) equations (2.1) are solved for the y's as functions of the z's and these expressions are substituted in (1.11), the resulting equations give the relations between the x's and the z's. In order that the latter be projective coördinates, it is necessary that

$$\left(\frac{\partial^2 x^i}{\partial y^j \partial y^k}\right)_0 = \left(\frac{\partial^2 x^i}{\partial z^j \partial z^k}\right)_0,$$

from which in consequence of (2.9) it follows that the solution φ must be such that

$$\left(\frac{\partial^2 z^i}{\partial y^j \partial y^k}\right)_0 = 0.$$

Since $(\varphi)_0 = 1$, we have from (2.1) that we must have $\left(\frac{\partial \varphi}{\partial y^i}\right)_0 = 0$. This condition is satisfied by $\bar{\varphi}$, defined by (2.7), and this is the only solution, because τ in (2.8) being any homogeneous function of first degree in the y's does not satisfy the condition $\left(\frac{\partial \tau}{\partial y^i}\right)_0 = 0$. Accordingly, if in (2.1) we use the function $\bar{\varphi}$, defined by (2.7), the coördinates z^i are such that

$$\left(\frac{\partial}{\partial z^i}\log\left|\frac{\partial x}{\partial z}\right|\right)_0 = 0, \tag{2.14}$$

as follows from (1.14). We call the z's, thus defined uniquely with respect to the x's, the *projective normal coördinates* of the space corresponding to the x's.

3. If we take another general coördinate system x'^i and define projective coördinates z'^i, with origin at P by means of the function $\bar{\varphi}'$ analogous to $\bar{\varphi}$ defined by (2.7), then in consequence of (2.9) and (2.13) we have

$$z'^i = \left(\frac{dx'^i}{dz'}\right)z' = \left(\frac{\partial x'^i}{\partial x^j}\frac{dx^j}{dz}\frac{dz}{dz'}\right)_0 z' = a^i_j z^j \left(\frac{dz}{dz'}\right)_0 \frac{z'}{z},$$

where the constants a^i_j are defined by

$$\left(\frac{\partial x'^i}{\partial x^j}\right)_0 = a^i_j. \tag{3.1}$$

Between the projective parameters z and z' for the respective coördinate systems z^i and z'^i and the jacobian of the transformation of these coordinates the following relation holds:[7]

$$\frac{dz'}{dz} = \left|\frac{\partial z'}{\partial z}\right|^{\frac{2}{n+1}}. \tag{3.2}$$

In consequence of (2.9)

$$\left(\left|\frac{\partial z'}{\partial z}\right|\right)_0 = \left(\left|\frac{\partial x'}{\partial x}\right|\right)_0 = |a^i_j| \equiv A.$$

Consequently we have

$$z'^i = a^i_j z^j \frac{z'}{z} A^{-\frac{2}{n+1}}.$$

Since the relation between the z's and z''s is independent of the paths, we must have

$$\frac{z'}{z} = \frac{A^{\frac{2}{n+1}}}{f} \tag{3.3}$$

where f is a determinate function of the z's. Differentiating this equation we have

$$\frac{dz'}{dz} = \frac{1}{f^2}\left(f - \frac{\partial f}{\partial z^j} z^j\right) A^{\frac{2}{n+1}}. \tag{3.4}$$

From the above equations follow the equations

$$z'^i = \frac{a^i_j z^j}{f}. \tag{3.5}$$

The jacobian of this transformation is[3]

$$\left|\frac{\partial z'}{\partial z}\right| = \frac{1}{f^{n+1}}\left(f - \frac{\partial f}{\partial z^j} z^j\right) A. \tag{3.6}$$

Then from (3.2), (3.4) and (3.6) we have

$$f - \frac{\partial f}{\partial z^j} z^j = 1,$$

of which the general solution is $1 + \sigma$, where σ is a homogeneous function of the first degree in the coördinates z^i. Consequently (3.6) reduces to

$$\left|\frac{\partial z'}{\partial z}\right| = \frac{1}{(1+\sigma)^{n+1}} A. \tag{3.7}$$

In consequence of (2.14), (2.9) and similar equations in the x''s and z''s, of (3.7), and the law of multiplication of jacobians we have

$$\left(\frac{\partial}{\partial x^i} \log\left|\frac{\partial x'}{\partial x}\right|\right)_0 = \left(\frac{\partial}{\partial z^i} \log\left|\frac{\partial z'}{\partial z}\right|\right)_0 = -(n+1)\left(\frac{\partial \sigma}{\partial z^i}\right)_0.$$

If then from the transformation of the x's and x''s we have constants a_i defined by

$$\left(\frac{\partial}{\partial x^i} \log\left|\frac{\partial x'}{\partial x}\right|\right)_0 = -a_i(n+1), \tag{3.8}$$

we obtain $\left(\frac{\partial \sigma}{\partial z^i}\right)_0 = a_i$. Put $\sigma = a_i z + \rho$, where ρ is a homogeneous function of the first degree. Since we must have $\left(\frac{\partial \rho}{\partial z^i}\right)_0 = 0$, it follows that $\rho = 0$. Consequently the relation between the z's and z''s is the linear fractional form

$$z'^i = \frac{a^i_j z^j}{1 + a_j z^j}, \tag{3.9}$$

where a^i_j and a_i are defined by (3.1) and (3.8), respectively.

4. In §2 it was seen that by taking for φ any solution of equation (2.5) the coördinates z^i defined by (2.1) satisfy (2.6), (2.12) and (2.13). We shall show that there is a solution φ' of the equation in the y''s analogous to (2.5) for which the coördinates z'^i are given by (3.9).

If we proceed with the equations $y'^i = \left(\frac{dx'^i}{dx'}\right)_0 x'$ in a manner similar to

that which led to (3.3), (3.4), (3.5) and (3.6), we get

$$\begin{aligned} \frac{x'}{x} &= \frac{A^{\frac{2}{n+1}}}{\rho}, \quad \frac{dx'}{dx} = \frac{1}{\rho^2}\left(\rho - \frac{\partial \rho}{\partial y^i} y^i\right) A^{\frac{2}{n+1}}, \\ y'^i &= \frac{a^i_j y^j}{\rho}, \quad \left|\frac{\partial y'}{\partial y}\right| = \frac{1}{\rho^{n+1}}\left(\rho - \frac{\partial \rho}{\partial y^i} y^i\right) A, \end{aligned} \tag{4.1}$$

where ρ is a function of the y's. Moreover, since x and x' are the projective parameters for the x's and x''s, we have[7]

$$\frac{dx'}{dx} = \left|\frac{\partial x'}{\partial x}\right|^{\frac{2}{n+1}}. \tag{4.2}$$

If we put

$$\varphi' = \frac{\varphi + a_j y^j}{a^i_j y^j} y'^i \qquad (i \text{ not summed}), \tag{4.3}$$

where a_j are any constants, because of (4.1) we have

$$\varphi = - a_j y^j + \rho \varphi'.$$

In consequence of (4.1) we have

$$\begin{aligned} \frac{\partial \varphi}{\partial y^i} y^i &= - a^i y^i + \frac{\partial \rho}{\partial y^i} \cdot y^i \cdot \varphi' + \frac{\partial \varphi'}{\partial y'^k} \frac{\partial y'^k}{\partial y^i} y^i \\ &= \varphi - \left(\rho - \frac{\partial \rho}{\partial y^i} y^i\right)\left(\varphi' - \frac{\partial \varphi'}{\partial y'^j} y'^j\right). \end{aligned} \tag{4.4}$$

From (4.1) and (4.2) we have

$$\left|\frac{\partial x'}{\partial x}\right| = \frac{1}{\rho^{n+1}} A \left(\rho - \frac{\partial \rho}{\partial y^i} y^i\right), \qquad \left|\frac{\partial x'}{\partial x}\right| \cdot \left|\frac{\partial y}{\partial y'}\right| = \left(\rho - \frac{\partial \rho}{\partial y^i} y^i\right)^{\frac{n-1}{2}}.$$

Hence from (4.4) and (2.5) we have

$$\varphi' - \frac{\partial \varphi'}{\partial y'^j} y'^j = \left|\frac{\partial x'}{\partial y'}\right|^{\frac{2}{n-1}},$$

which is the analogous equation in the y''s. Using φ' defined by (4.3) in the definition of the z'^i, we have the equations (3.9). Thus we have general sets of coördinates satisfying (3.9) where a_j are arbitrary constants. But these are projective normal coördinates only when $\bar{\varphi}$, defined by (2.7), and the analogous $\bar{\varphi}'$ are used for the definition of z^i and z'^i, and the constants a_j are given by (3.8).

[1] "On the Projective and Equiprojective Geometries of Paths," *Proc. Nat. Acad. Sci.*, **11**, pp. 199–203, 1925.

[2] "Projective Normal Coördinates for the Geometry of Paths," *Ibid.*, **11**, pp. 204–207, 1925.

[3] "Non-Riemannian Geometry," *Am. Math. Soc. Colloq. Publica.*, **8**, 1927. Hereafter a reference to this book is of the form E §45.

[4] E. §§22, 40. Formula (22.8) is not correct so as to insure that the quantities are symmetric in all the indices; the form to insure this requirement is readily obtained.

[5] E., p. 111.

[6] E., §41.

[7] E., p. 107.

[8] E., p. 113.

ON THE NORM-RESIDUE SYMBOL IN THE THEORY OF CYCLOTOMIC FIELDS

By H. S. Vandiver

Department of Mathematics, University of Texas

Communicated September 18, 1930

Let l be an odd prime and $\zeta = e^{2i\pi/l}$; $\mathfrak{l} = (1 - \zeta)$; k the field defined by ζ and ω and θ integers in k prime to $\mathfrak{l}$. Then one way of expressing the norm residue symbol for the field k is

$$\left\{\frac{\omega, \theta}{\mathfrak{l}}\right\} = \zeta^P,$$

where

$$P \equiv -l_1(\theta)\,\frac{N(\omega) - 1}{l} + l_1(\omega)\,\frac{N(\theta) - 1}{l} + \sum_{s=2}^{l-2} (-1)^{s-1}\, l_s(\theta)\, l_{l-s}(\omega) \pmod{l}; \tag{1}$$

$$\theta = a_0 + a_1\,\zeta + \ldots + a_{l-2}\,\zeta^{l-2},$$

the a's rational integers;

$$\theta(e^v) = a_0 + a_1\,e^v + a_2\,e^{2v} + \ldots + a_{l-2}\,e^{(l-2)v};$$

$$l_s(\theta) = \left[\frac{d^s \log \theta(e^v)}{dv^s}\right]_{v=0};$$

$N(\theta)$ is the norm of θ; with similar definitions for ω. This symbol has been treated by a number of writers.[1] So far, apparently, the value of P has always been represented by summation of the type mentioned above or analogous to it. In the present note I shall give a method for carrying out this summation. To effect this we expand the expression

$$G = \frac{d^n}{dv^n}\left(\theta(e^v)^l \cdot \frac{d \log \theta(e^v)}{dv}\right)$$

considered as a product, by Leibnitz's theorem. Since

$$d(\theta(e^v)^l) = l\theta(e^v)^{l-1}\, d\theta$$

and

$$\frac{d \log \theta(e^v)}{dv} = \frac{1}{\theta(e^v)} \frac{d(\theta(e^v))}{dv},$$

we find

$$G = (\theta(e^v))^l \frac{d^n}{dv^n}\left(\frac{1}{\theta(e^v)} \cdot \frac{d(\theta(e^v))}{dv}\right) + \frac{lA(e^v)}{B(e^v)}, \qquad (2)$$

where A is a polynomial in e^v with rational integral coefficients and $B(e^v)$ is a power of $\theta(e^v)$ and hence is also a polynomial of the same type as A.

If $\theta(1) \equiv a \pmod{l}$, a must be prime to l, and we therefore have from (2) by setting $v = 0$,

$$G_{v=0} \equiv a^l \left[\frac{d^n}{dv^n}\left(\frac{1}{\theta(e^v)} \frac{d(\theta(e^v))}{dv}\right)\right]_{v=0} \pmod{l},$$

$$\equiv a^l\, l_{n+1}(\theta) \pmod{l}.$$

But also from the definition of G we have

$$G_{v=0} = \left[\frac{d^n}{dv^n}\left(\theta(e^v)^{l-1} \frac{d(\theta(e^v))}{dv}\right)\right]_{v=0}.$$

We may set

$$\theta(e^v)^{l-1} \frac{d(\theta(e^v))}{dv} = c_0' + c_1' e^v + \ldots + c_l' e^{vl}$$

where the (c')'s are rational integers. Similarly

$$(\omega(e^v))^{l-1} \frac{d(\omega(e^v))}{dv} = d_0' + d_1' e^v + \ldots + d_h' e^{vh}.$$

Hence

$$a^l\, l_s(\theta) = c_1' + 2^{s-1} c_2' + 3^{s-1} c_3' + \ldots + l^{s-1} c_l'$$

and if $\omega(1) \equiv b \pmod{l}$, then b is prime to l and

$$b^l\, l_s(\omega) = d_1' + 2^{s-1} d_2' + \ldots + h^{s-1} d_h'.$$

Now use the relation

$$c_{rl+i}(rl + i)^{s-1} \equiv i^{s-1}\, c_{rl+i} \pmod{l},$$

and collect the terms in the last two relations, we find

$$a^l\, l_s(\theta) \equiv c_1 + 2^{s-1}\, c_2 + \ldots + (l-1)^{s-1}\, c_{l-1} \pmod{l},$$

and

$$b^l\, l_s(\omega) \equiv d_1 + 2^{s-1}\, d_2 + \ldots + (l-1)^{s-1}\, d_{l-1} \pmod{l},$$

where the c's and d's are rational integers. Consider the sum

$$\sum_{s=1}^{l-1} (-1)^{s-1}\, a^l\, b^l\, l_s(\theta) l_{l-s}(\omega) = S.$$

Using the above congruences, this may be written as

$$a^l\, b^l \sum_{i,j} c_i\, d_j (i^{l-2} - i^{l-3}\, j + \ldots - j^{l-2})$$

where i and j range independently over the integers $1, 2, 3, \ldots, l-1$. The term involving the summation is congruent modulo l to

$$\sum_{i \neq l-j} \frac{i^{l-1} - j^{l-1}}{i+j} + (l-1)(c_1\, d_{l-1} + c_2\, d_{l-2} + \ldots + c_{l-1}\, d_1).$$

Since $i^{l-1} \equiv j^{l-1} \equiv 1 \pmod{l}$, we have

$$S \equiv -\, a^l\, b^l \sum_{s=1}^{l-1} c_s\, d_{l-s} \pmod{l}.$$

Applying this to (1) we obtain P in the form desired.

For the case where θ and ω are each integers in the field which may be expressed in the binomial form

$$\theta = c_0 + \zeta c_1 \quad \text{and} \quad \omega = d_0 + \zeta d_1,$$

we may employ the modification of the above-mentioned method and obtain a particularly simple form for P. In this case, we use

$$H = \frac{d^n}{dv^n}\left(\left(c_0^l + e^{vl}\, c_1^l\right) \frac{e^v\, c_1}{c_0 + e^v\, c_1}\right)$$

in lieu of G and note that

$$\theta(1) l_{n+1}(\theta) \equiv H_{v=0} \equiv \theta(1)\left[\frac{d^n}{dv^n}\left(\frac{e^v\, c_1}{c_0 + e^v\, c_1}\right)\right]_{v=0} \pmod{l},$$

and also

$$H_{v=0} \equiv -\, c_0^{l-2}\, c_1 + 2^n\, c_0^{l-3}\, c_1^2 - 3^n\, c_0^{l-4}\, c_1^3 + \ldots + (l-1)^n\, c_1^{l-1} \pmod{l},$$

and we see that $H_{v=0}$ is completely determined without the necessity of raising $\theta(e^v)$ to the power $l-1$, and proceeding as in the general case

P is obtained in convenient form. This special method may be applied also to the two forms

$$\theta = m + n\zeta^i; \qquad \omega = h + g\zeta^j;$$

where i and j are arbitrary integers, but the result for P is more complicated than in the case $i = j = 1$.

[1] Kummer, Hilbert, Bericht, Sitzungsberichte D. M. V., 1894, 413; Takagi, Crelle, 157: 230–8 (1927); Hasse, numerous papers in Crelle, 153–155. The form used here was employed by the author in several papers, for example, in the *Ann. Math.* [2], 30, 553 (1930). It follows easily from that given by Hilbert (loc. cit.).

ON THE SECOND FACTOR OF THE CLASS NUMBER OF A CYCLOTOMIC FIELD

By H. S. Vandiver

Department of Mathematics, University of Texas

Communicated September 18, 1930

The second factor of the class number of a cyclotomic field defined by l, an odd prime, is a rational integer which is generally expressed as the quotient of two determinants.[1] It is known that for all values of $l < 23$ this second factor is unity; that it is not unity for all values of l was shown by Kummer, who found that in some cases it is divisible by 2 or 3.

Many questions in the theory of cyclotomic fields depend on the divisibility of this second factor by l; it is known[2] that the second factor of the class number is divisible by l only if the first factor is divisible by l. This condition is not sufficient,[3] however. In the present paper we shall prove the following

THEOREM. *A necessary and sufficient condition that the second factor of the class number of the field defined by* $\zeta = e^{2i\pi/l}$ *be divisible by* l *is that at least one of the units*

$$E_i; \quad i = 1, 2, \ldots, l_1; \quad l_1 = (l-3)/2$$

be the l*th power of a unit in* $k(\zeta)$. *Here*

$$E_i = \epsilon^R; \quad R = r^{l^2} + sr^{l^2-2i} + s^2 r^{l^2-4i} + \ldots + s^{l_1} r^{l^2-2il};$$

$$\epsilon = \left(\frac{(1-\zeta^r)(1-\zeta^{-r})}{(1-\zeta)(1-\zeta^{-1})}\right)^{1/2};$$

r is a primitive root of l and we are employing the Kronecker-Hilbert notation of symbolic powers, the substitution s denoting (ζ/ζ^r).

That the condition is necessary is proved in two other papers[4] on this subject. Kummer[5] stated without proof that in the special case where the first factor of the class number is divisible by l, but not by l^2 that the condition stated in the theorem is also sufficient. The methods employed in his article referred to readily yields a proof of this sufficiency, but the problem is far more complicated in the general case and other methods appear necessary.

To prove it sufficient in the general case, we employ the expression for the second factor of the class number that was used by Miss Stafford and myself in our paper already cited. It is

$$\frac{n_1 n_2 \ldots n_{l_1}}{\Delta}$$

where Δ is the determinant

$$\begin{vmatrix} b_{11} & b_{21} & \ldots & b_{l_1 1} \\ b_{12} & b_{22} & \ldots & b_{l_1 2} \\ \ldots & \ldots & & \\ \ldots & \ldots & & \\ b_{1l_1} & b_{2l_1} & \ldots & b_{l_1 l_1} \end{vmatrix}$$

the b's and n's being rational integers such that

$$\gamma_s^{n_s} = \epsilon_1^{b_{1s}} \epsilon_2^{b_{2s}} \ldots \epsilon_{l_1}^{b_{l_1 s}}; \qquad (1)$$

$$s = 1, 2, \ldots l_1.$$

In our paper cited, it was stated that there was no common factor of all the integers n_s, b_{1s}, $b_{2s} \ldots b_{l_1 s}$. However, it will be more convenient here to assume only that l is not a common factor of all these integers. The derivation of the above expression for the second factor of the class number in our article justifies this assumption. Also, in (1) the unit ϵ_i is the unit obtained from ϵ by the substitution $(\zeta/\zeta^{r^{i-1}})$. In addition, we may assume that all the b's in any of the relations (1) cannot be divisible by l for it would follow from this that the corresponding n is divisible by l since we shall define the γ's as forming a fundamental system of units in our field (as yet undetermined).

Let $\eta_1, \eta_2, \ldots, \eta_{l_1}$ be any fundamental system of units in the field; then, since the E's are known to form an independent system, we may write, using symbolic powers, the a's, k's and c's being rational integers, $c_i \geqq 0$,

$$\eta_i^K = E^A; \quad K = k_i f^{i}; \quad A = a_{1i} + a_{2i}\, s + \ldots + a_{l_1 i}\, s^{l_1 - 1}; \qquad (1a)$$

$\overline{i} = 1, 2, \ldots, l_1$ and s, used in an exponent, stands for the substitution (ζ/ζ^r), when the k's are prime to l and not all the a's in a particular relation have the common factor l. Let $c > c_s$ and assume that the γ's in (1) form a set of fundamental units which have the property[6]

$$\gamma_i^{s-h_i} = \beta_i^{l^c}; \quad h_i = r^{2il^{c-1}}; \quad i = 1, 2, \ldots, l_1, \tag{2}$$

where β_i is a unit in $k(\zeta)$. We may then write, if the (b')'s, j's and d's are rational integers,

$$\gamma_i^J = E_1^{b'_{1i}} E_2^{b'_{2i}} \ldots E_{l_1}^{b'_{l_1 i}}; \quad J = j_i l^{d_i};$$

$$i = 1, 2, \ldots l_1,$$

where the j's are prime to l and not all the (b')'s in a particular equation are divisible by l. Raise each one of these equations to the power $k = k_1 k_2 \ldots k_{l_1}$, we obtain

$$\gamma_i^{M_i} = E_1^{e_{1i}} E_2^{e_{2i}} \ldots E_{l_1}^{e_{l_1 i}}; \quad m_i l^{d_i} = M_i; \tag{3}$$

$$i = 1, 2, \ldots, l_1,$$

where not all the exponents of the E's in a particular equation are divisible by l. Take the particular relation corresponding to $i = v$ then making the substitution s, and noting that $E_i^{s-r^{2i}} = \omega_i^l$ we have from (2) after noting also that ω_i is a power of one of the ϵ's,

$$\beta_v^{m_v l^{c+d_v}} \gamma_v^{V_1} = E_1^{r^2 e_{1v}} E_2^{r^4 e_{2v}} \ldots E_{l_1}^{r^{2l_1} e_{l_1 v}} \delta_1^l; \tag{3a}$$

$V = m_v l^{d_v} r^{2vl^{c-1}}$; the e's being rational integers and δ_1 a unit in $k(\zeta)$, which is the product of integral powers of the ϵ's.

The unit β_v may be written in the form

$$\eta_1^{x_1} \eta_2^{x_2} \ldots \eta_{l_1}^{x_{l_1}}$$

where the x's are rational integers. But

$$\eta_f^{l^c m_v}; \quad f = 1, 2, \ldots, l_1,$$

by $(1a)$ may be expressed as a product of l-th powers of the E's since $c > c_i$ and m_v is divisible by k. Hence it may be expressed as a product of l-th powers of the ϵ's if we use the relation

$$\epsilon_{l_1+1} = \epsilon_1^{-1} \epsilon_2^{-1} \ldots \epsilon_{l_1}^{-1}. \tag{4}$$

Employing this in $(3a)$ and substituting $\zeta^r, \zeta^{r^2}, \ldots, \zeta^{r^{l_1}}$, in turn in the same equation we have with (3)

$$\gamma_v^{V_t} = E_1^{r^{2t} e_{1v}} E_2^{r^{4t} e_{2v}} \ldots E_{l_1}^{r^{2l_1 t} e_{l_1 v}} \delta_t^l;$$

$$V_t = m_v l^{d_v} r^{2vtl^{c-1}}; \quad t = 0, 1, \ldots, l_1;$$

and δ_t is the product of integral powers of the ϵ's. Now suppose that $w \neq v$ and raise each member of the last equation to the power r^{-2wt}

and multiply together the resulting equations, obtained by allowing t to range as mentioned, as they stand. The exponent of γ_v is

$$m_v l^{d_v} \sum_{t=0}^{l_1} r^{(2vl^c-1-2w)t} = \frac{r^{(l-1)(vl^c-1-1)} - 1}{r^{2vl^c-1-2w} - 1},$$

and since $2vl^{c-1} \not\equiv 2w \pmod{l-1}$ we infer that this exponent is a multiple of l. Similarly the exponents of the E's are each divisible by l except that of E_v. Hence we have for $w \neq v$

$$E_w^{e_{wv} \cdot \frac{l-1}{2}} \theta_1^l = 1, \quad \text{or} \quad E_w^{e_{wv}} = \theta^l.$$

where θ is a product of integral powers of the ϵ's. Substituting in (3) we have, if $i = 1, 2, \ldots, l_1$,

$$\gamma_i^{M_i} = E_i^{e_{ii}} \omega^l; \quad M_i = m_i l^{d_i},$$

where ω is the product of integral powers of the ϵ's. The relation (3) can now be written

$$\gamma_i^{M_i} = E_i^{e_{ii}} \lambda_i^l. \tag{5}$$

Raise each side of (5) to the power r^{2il_1}, and in the resulting relation substitute for ϵ_{l_1+1} its value as given in (4). The expression on the right is then expressed entirely as products of powers of the ϵ's and this shows that in (1) we may take

$$b_{ii} \equiv e_{ii} r^{l^2} (r^{2i(l_1-l+1)} - 1) \pmod{l}.$$

Hence the determinant $\Delta \equiv r^{l_1 l^2} \Delta' \prod_{i=1}^{l_1} e_{ii} \pmod{l}$ and Δ' is congruent modulo l to

$$\begin{vmatrix} N_{1,l_1} & N_{1,l_1-1} & \cdots & N_{1,1} \\ N_{2,l_1} & N_{2,l_1-1} & \cdots & N_{2,1} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ N_{l_1,l_1} & N_{l_1,l_1-1} & \cdots & N_{l_1,1} \end{vmatrix}; \quad N_{i,t} = r^{2it} - 1.$$

In the above, add to the elements of the first row the corresponding elements of each of the other rows and we find, since

$$r^{2i} + r^{4i} + \ldots + r^{2il_1} = \frac{r^{i(l-1)} - 1}{r^{2i} - 1} - 1 \equiv -1 \pmod{l};$$

$$i = 1, 2, \ldots, l_1,$$

that Δ' is congruent modulo l to a determinant in which each element of the first row is equal to $(l-1)/2$. Set $\Delta' = (l-1)/2 \cdot \Delta''$. Now

in Δ'' add to the elements of each row the corresponding element of the first row and we find that it reduces to

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ (r^4)^{l_1} & (r^4)^{l_1-1} & \cdots & r^4 \\ (r^6)^{l_1} & (r^6)^{l_1-1} & \cdots & r^6 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ (r^{2l_1})^{l_1} & (r^{2l_1})^{l_1-1} & \cdots & r^{2l_1} \end{vmatrix}$$

and this is known to be the product of factors all of the type $r^a - r^b$; $a \neq b$, where a and b are each less than $l - 1$. Hence Δ'' and therefore Δ is prime to l.

With the use of these results we shall now prove the condition in our theorem sufficient. Assume that $E_t = \sigma^l$ where σ is a unit in $k(\zeta)$. We then have from (5)

$$\gamma_t^{M_t} = E_t^{e_{tt}} \lambda_t^l = \xi^l$$

where ξ is a unit in $k(\zeta)$. Since the γ's form a fundamental system of units, we may write, if the h's are integral,

$$\xi = \gamma_1^{h_1} \gamma_2^{h_2} \cdots \gamma_{l_1}^{h_{l_1}},$$

and substituting this expression for ξ in the previous relation we infer, since the γ's are independent,

$$M_t = h_t l,$$

and therefore from (1), $n_t = 0 \pmod{l}$, and since we expressed the second factor f of the class number in the form

$$f = \frac{n_1 n_2 \cdots n_{l_1}}{\Delta},$$

we have that $f \equiv 0 \pmod{l}$, since Δ is prime to l.

I shall now indicate, with few details, a second proof of the sufficiency of the condition in the theorem which is quite different from the first proof in character. Consider[7] the theorem VI given by Pollaczek. If we assume that there exists an i in the set 1, 2, 3, . . ., l_1 such that $E_i = \delta^l$ and also that the second factor of the class number of $k(\zeta)$ is prime to l, then there exists none of the $\mathfrak{q}$ ideal classes referred to in Pollaczek's theorem, hence, in the notation of his theorem, $z_{b_0} = 1$ for each of the non-quadratic residues b_0 such that $\mathfrak{p}^{s-b_0} \sim 1$. Also, the proof of his result shows that since there exists no $\mathfrak{q}$ class in the irregular class group, then all the singular primary numbers in $k(\zeta)$ are real. Now it will be shown in a forthcoming paper of the writer's[8] that if

$$\left[\frac{d^n \log \rho(e^v)}{dv^n}\right]_{v=0} \equiv 0 \pmod{l}.$$

It therefore follows that ρ is a singular primary number which is not real and we have a contradiction.

In the first part of this paper, I referred to the fact that $E_1, E_2, \ldots, E_{l_1}$ form an independent system of units in $k(\zeta)$. This result was used by Kummer but I have as yet been unable to find an explicit proof in his papers. The result follows by use of some of the methods in this paper, since if we assume that the above set is not independent, then

$$E_1^{a_1} E_2^{a_2} \ldots E_{l_1}^{a_{l_1}} = 1$$

where the a's are integers, not all $\equiv 0 \pmod{l}$. If we substitute for each E its value in terms of the ϵ's, then we obtain

$$\epsilon_1^{t_1} \epsilon_2^{t_2} \ldots \epsilon_{l_1}^{t_{l_1}} = 1$$

and where

$$t_i = a_1 L_{i1} + a_2 L_{i2} + \ldots + a_{l_1} L_{il_1}$$

$$i = 1, 2, \ldots, l_1;$$

$$L_{ik} = r^{l^2-2(i-1)k} - r^{l^2-2kl_1};$$

$$k = 1, 2, \ldots, l_1.$$

Now since the ϵ's form an independent system, we infer that each t is zero, but the determinant K of the coefficients of the a's is not divisible by l since we can multiply, for each k, the elements of the kth column by $r^{2kl_1-l^2}$ and this gives our former determinant Δ' which we have already proved prime to l. This shows that each a is divisible by l which is a contradiction.

[1] Report on the theory of algebraic numbers, *Bull. Nat. Res. Coun.*, **2**, February, 1928, p. 34; Fueter, *Synthetische Zahlentheorie*, Berlin and Leipzig, **1925**, second ed., p. 223.

[2] Report on Alg. Nos., loc. cit., **2**, pp. 37–40.

[3] Stafford and Vandiver, *Proc. Nat. Acad. Sci.*, **16**, 139–150 (1930).

[4] Stafford and Vandiver, loc. cit.; Vandiver, *Bull. Am. Math. Soc.*, **35**, 333–5 (1929).

[5] Abhandlungen der Koniglichen Akademie der Wissenschaften zu Berlin, *Math.-Phys. Klasse*, 1857.

[6] Pollaczek, *Math. Zeits.*, **20**, 9 (1924).

[7] Loc. cit., p. 22.

[8] *Trans. Amer. Math. Soc.*, 1930, "On the Composition of the Group of Ideal Classes in an Irregular Cyclotomic field."

[9] *Bull. Amer. Math. Soc.*, **25**, 458–461 (1919); *Ann. Math.* [2], **26**, 231–223 (1925).

[10] These PROCEEDINGS, **15**, 205 (1929).

PERIODIC RECURRING SERIES

BY E. T. BELL

DEPARTMENT OF MATHEMATICS, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated September 25, 1930

1. THEOREM. Let x be restricted to rational integral values, and let $a_1, \ldots, a_r$ be constant rational numbers not all zero. Let $f(x)$ be a solution of the linear difference equation

$$f(x + r) + a_1 f(x + r - 1) + \ldots + a_r f(x) = 0 \qquad (1.1)$$

of order r, whose characteristic equation

$$y^r + a_1 y^{r-1} + \ldots + a_r = 0 \qquad (1.2)$$

is irreducible (in the rational domain). *Then $f(x)$ has the proper period m, > 0, which is unique,*

$$f(x + m) = f(x),\ f(x + n) \neq f(x)\ (n \neq hm,\ h \text{ integer}) \qquad (1.3)$$

when and only when $r = \phi(m)$ (the number of integers $\leq m$ and prime to m) *and the roots of the equation* (1.2) *are the primitive mth roots of unity.*

As the proof is short it is given in §3. For the moment we note the following.

COROLLARY. *When $f(x)$ has the proper period m, the coefficients $a_1, \ldots, a_r$ are integers.*

COROLLARY. *When $f(x)$ has the proper period $2m$, $f(x + m) = -f(x)$.*

COROLLARY. *If $f(x)$ has the proper period $m > 2$, the order r is even,* (*since $\phi(m)$ is even when $m > 2$*). The only periodic recurring series of odd order are defined by

$$f(x + 1) - f(x) = 0,\ f(x + 1) + f(x) = 0,$$

with the respective periods 1,2.

COROLLARY. *If $f(x)$ has the proper period m, the characteristic equation* (1.2) *is*

$$\Pi\,(y^a - 1)^{\mu(b)} = 0, \qquad (1.4)$$

where μ is Möbius' (or Mertens') function, and the product refers to all pairs (a,b) of positive divisors of m such that $m = ab$.

Let $f_0(x), \ldots, f_{r-1}(x)$ be the linearly independent solutions of (1.1) determined by the initial conditions $f_i(j) = \delta_{ij}$ $(i,j = 0, \ldots, r - 1)$, where $\delta_{ii} = 1$, $\delta_{ij} = 0$, $i \neq j$, and let $f(0), \ldots, f(r - 1)$ be arbitrary constants. Then the general solution of (1.1) is

$$f(x) = f(0)\, f_0(x) + \ldots + f(r - 1) f_{r-1}(x). \qquad (1.5)$$

COROLLARY. *By* (1.5) *there are* $\infty^{\phi(m)}$ *recurring series having the proper period* m. *The terms of such a series are integers when and only when* $f(j)$ $(j = 0, \ldots, \phi(m) - 1)$ *are integers.*

2. *Remarks on Distribution.*—The values of the possible orders r and periods m raise some interesting questions in distribution. It has been noted that there are precisely two equations (1.1) whose solutions are periodic recurring series of odd order ($r = 1$). All cases are included in the following.

COROLLARY. *If* $\phi^{-1}(r)$ *denotes the (necessarily finite) number of integers whose* ϕ*-function has the constant value* r, *there exist precisely* $\phi^{-1}(r)$ *equations* (1.1) *whose solutions are periodic recurring series of order* r; *the periods of these series are the solutions* m *of* $\phi(m) = r$.

From the table (p. 395) in Lucas' *Théorie des Nombres*, we see therefore that the only $2r \leq 100$ for which no such equation of order $2r$ exists are the 13 numbers

$$14, 26, 34, 38, 50, 62, 68, 74, 76, 86, 90, 94, 98.$$

In the same range the maximum number of such equations of order $2r$ occurs for $r = 36, 48$, and is 17; the minimum number 2 occurs for even orders precisely 12 times, the largest order being 78; the longest period is 420, for the order 96. The distribution of the values of $\phi^{-1}(r)$ is extremely irregular. It is possible, however, to express $\phi^{-1}(r)$ in terms of known arithmetical functions; this will be done in another note. The total number of equations (1.1) of orders ≤ 100 having periodic solutions is 198.

Reuschle's *Tafeln complexer Prinzahlen welche aus Wurzeln der Einheit gebildet sind* (Berlin, 1875), contains incidentally (pp. 193–440, 467–671) all the equations (1.4) worked out fully for odd composite $m \leq 105$ and for $m = 4n$, n composite, $3 \leq n \leq 30$. It is interesting to notice that the only one of these equations (in the ranges indicated) having a coefficient other than 0,1 or -1 occurs when $m = 105$, when the coefficient -2 appears once.

3. *Proof.*—Let $y_1, \ldots, y_r$ be the roots of the equation (1.2). Then any solution of (1.1) is of the form

$$f(x) = c_1 y_1^x + \ldots + c_r y_r^x,$$

where the c's are constants. If this $f(x)$ has the period m, $f(x + m) = f(x)$, and

$$f(x) = c_1 y_1^{x+m} + \ldots + c_r y_r^{x+m}.$$

Hence

$$c_1(y_1^m - 1)y_1^x + \ldots + c_r(y_r^m - 1)y_r^x = 0.$$

In this take $x = 0, \ldots, r - 1$. The Vandermonde determinant $D(y_1,$

$\ldots, y_r) \neq 0$, since $y_1, \ldots, y_r$ are distinct by the irreducibility of (1.2). Hence the set of r homogeneous equations just constructed has the unique set of solutions

$$c_1(y_1^m - 1) = 0, \ldots, c_r(y_r^m - 1) = 0. \tag{3.1}$$

If therefore none of $c_1, \ldots, c_r$ vanish, $y_1, \ldots, y_r$ are mth roots of unity. We shall show that the hypothesis that precisely h of $c_1, \ldots, c_r$ vanish, where $0 < h \leq r$, leads to a contradiction. For, if precisely h vanish, $f(x)$ is a solution of an equation of type (1.1) of order $r - h$. By hypothesis the characteristic equation of this equation is irreducible and has rational coefficients. But it has $r - h$ roots in common with (1.2), which is irreducible. Since $r - h < r$, we have a contradiction. Hence $y_1, \ldots, y_r$ are mth roots of unity.

It remains to be proved that $y_1, \ldots, y_r$ are primitive. Let $y^m - 1 = P_0(y) \ldots P_s(y)$ be the resolution of $y^m - 1$ into factors irreducible in the rational domain, and let the roots of $P_0(y) = 0$ be the $\phi(m)$ primitive mth roots of unity. Then the roots of $P_1(y) = 0, \ldots, P_s(y) = 0$ are imprimitive mth roots of unity. But since m is by hypothesis a proper period, m is the least positive integer for which the equations (3.1) hold. Hence $y_1, \ldots, y_r$ are not nth roots of unity, $n < m$, and therefore they are the roots of $P_0(y) = 0$.

NOTE ON SINGULARITIES OF POWER SERIES

BY H. F. BOHNENBLUST

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated October 3, 1930

THEOREM: If

(1) the power series $\sum_{n=0}^{\infty} a_n z^n$, $\sum_{n=0}^{\infty} p_n z^n$, and $\sum_{n=0}^{\infty} a_n p_n z^n$ have radii of convergence equal to one; and

(2) $\sum_{n=0}^{\infty} a_n z^n$ has on its closed circle of convergence a single singularity at the point $z = 1$; and

(3) the coefficients p_n are non-negative, $p_n \geqq 0$;

then the power series $\sum_{n=0}^{\infty} a_n p_n z^n$ is necessarily singular at the point $z = 1$.

This theorem is on one hand a generalization of the theorem of Pringsheim on singularities of power series with non-negative coefficients. (Pringsheim's theorem is obtained by putting all $a_n = 1$.) On the other

hand it is a contribution to the Hadamard theorem on singularities of the Hadamard product of two power series. Hadamard theorem applied to $\sum_{n=0}^{\infty} a_n z^n$ and $\sum_{n=0}^{\infty} p_n z^n$ states that the point $z = 1$ is a possible singular point of $\sum_{n=0}^{\infty} a_n p_n z^n$, while our theorem states that this point is necessarily singular.

The proof is an immediate consequence of the theorems of Pringsheim and Hadamard. Considering the first condition of the theorem, we obtain that to any given positive number ϵ, there exists a number $N(\epsilon)$, such that

$$\left.\begin{aligned} |a_n| &< (1+\epsilon)^n \\ p_n &< (1+\epsilon)^n \end{aligned}\right\} \text{for all } n \geqq N(\epsilon),$$

and $|a_n| p_n > (1-\epsilon)^n$ for infinitely many indices. A simple calculation shows that, as soon as $\epsilon < 1$, we have

$$|a_n|^2 p_n < (1+7\epsilon)^n \text{ for all } n \geqq N(\epsilon),$$

and $|a_n|^2 p_n > (1-3\epsilon)^n$ for infinitely many indices. These inequalities express that the radius of convergence of the power series $\sum_{n=0}^{\infty} |a_n|^2 p_n z^n$ is equal to one. All the coefficients being positive we can apply the theorem of Pringsheim, which shows that the point $z = 1$ is singular for that series.

But the same series is the Hadamard product of $\sum_{n=0}^{\infty} a_n p_n z^n$ and $\sum_{n=0}^{\infty} \bar{a}_n z^n$. The latter one has (condition 2) a single singularity at $z = 1$ on its closed circle of convergence. Assuming (against the theorem) $\sum_{n=0}^{\infty} a_n p_n z^n$ to be regular at $z = 1$, it follows that $\sum_{n=0}^{\infty} |a_n|^2 p_n z^n$ should be regular at the same point, contradicting the above result. This contradiction proves the theorem.

It is essential to suppose that the power series $\sum_{n=0}^{\infty} a_n z^n$ has only one singularity on its closed circle of convergence; for, given any arc on the unit circle, there exists a function, singular at $z = 1$ and at a point $z_0 (z_0 \neq 1)$ of the given arc, which when "multiplied" by an adequate power series $\sum_{n=0}^{\infty} p_n z^n$, satisfying conditions 1° and 3°, gives a power series regular at $z = 1$. In fact, on any arc there exists a point z_0, for which $z_0^l = -1$ (l being a positive integer). Let us then consider the function

$$\frac{1}{1-z} + \frac{2}{1-z\bar{z}_0} \sum_{n=0}^{\infty} a_n z^n,$$

where $a_n = 1 + 2\bar{z}_0^n$. On the other hand let us take

$$p_n = \begin{cases} 1/3 \text{ when } n = l.m,\ m \text{ even}, \\ 1 \text{ when } n = l.m,\ m \text{ odd}, \\ O \text{ otherwise}. \end{cases}$$

The "product" is the function $\sum_{n=0}^{\infty} C_n z^n$, where

$$C_n = \begin{cases} 1 \text{ when } n = l.m,\ m \text{ even}, \\ -1 \text{ when } n = l.m,\ m \text{ odd}, \\ O \text{ otherwise}. \end{cases}$$

This function is equal to $\dfrac{1}{1+z^l}$, and is therefore regular at $z = 1$.

Putting the coefficients p_n equal to one or to zero, we obtain as a corollary to our theorem:

"A power series having only one singular point z_0 on its closed circle of convergence is such that all its sub-series having the same circle of convergence are singular at z_0."

A METHOD OF OBTAINING NORMAL REPRESENTATIONS FOR A PROJECTIVE CONNECTION

By J. H. C. Whitehead

Department of Mathematics, Princeton University

Communicated October 3, 1930

Normal representations for a projective connection, Π, are characterized by the equations

$$P^{\alpha}_{jk} z^j z^k = 0, \tag{1.1}$$

where $P^{\alpha}_{\beta\gamma}$ are the components of Π in the normal representation $z + z^0$; (Greek letters, used as indices, will take the values $o, \ldots n$, and Roman letters the values $1, \ldots n$). In a paper in a forthcoming issue of the *Annals of Mathematics* we give a construction for obtaining such representations. In the introduction to that paper we give a brief historical account of generalized projective geometry, in which references to the literature may be found.

In this note we give an alternative method depending on solutions to the partial differential equations.

$$\left(\frac{\partial^2\theta}{\partial x^j \partial x^k} - \frac{\partial\theta}{\partial x^i}\Pi^i_{jk} - \theta\,\Pi^0_{jk}\right) X^j X^k = 0, \tag{1.2}$$

where $\Pi^\alpha_{\beta\gamma}$ are the components of the connection Π in the representation $x + x^0$, and X is an affine vector defined as follows. Let y^i be affine normal coördinates for Π^i_{jk}—treated as the components of an affine connection—the coördinate system x, and a given point q. Then X is to be the affine vector whose components in y are y^i. At any point near q, therefore, X touches the path joining that point to q.

If we put $Z = e^{x^0}\theta$, the equations (1.2) may be written

$$Z_{;\beta;\gamma}\, V^\beta V^\gamma = 0, \tag{1.3}$$

where the semicolon denotes projective differentiation with respect to Π, and V is any projective vector such that $V^i = X^i$. The identities $Z_{;\beta;0} = Z_{;0;\gamma} = 0$ arise out of the conditions $\Pi^\alpha_{\beta 0} = \delta^\alpha_\beta$. The equations (1.3) are invariant under all changes of representation, and so any solution Z is a projective scalar of index unity. We shall show that these equations admit a unique solution $Z = e^{x^0}\theta(x)$, where θ depends only on $x^1, \ldots x^n$, satisfying the initial conditions

$$\left(\frac{\partial Z}{\partial x^\beta}\right)_q = a_\beta,$$

where a_β are any given constants, and q a given point, together with a value, q^0, of the factor.

We take the $(n + 1)$ solutions Z^α determined by the conditions

$$\left(\frac{\partial Z^\alpha}{\partial x^\beta}\right)_q = \delta^\alpha_\beta, \tag{1.4}$$

and show that the normal representation for q and $x + x^0$ is given by

$$\begin{cases} z^0 = x^0 - q^0 + \log\theta^0 \\ z^i = \dfrac{\theta^i}{\theta^0}. \end{cases} \tag{1.5}$$

We shall do no more than establish this result, as a geometrical discussion of these representations, with further developments of the theory, are given in the paper to which we have referred.

2. We shall need the following lemma.

Let $A_{jk} = A_{kj}$ *be* $\dfrac{n(n+1)}{2}$ *analytic functions of* $x^1 \ldots x^n$, *which can be expanded about* $x = 0$ *in power series, convergent for* $|x^i| \leq \delta$, *and let* $a_0 \ldots a_\alpha$ *be arbitrary constants. Then the equation*

$$\left(\frac{\partial^2 y}{\partial x^j \partial x^k} - y\, A_{jk}\right) x^j x^k = 0 \tag{2.1}$$

will admit a solution $y(x)$, *regular near* $x = 0$, *which is uniquely determined by the initial conditions*

$$y(0) = a_0, \left(\frac{\partial y}{\partial x^i}\right)_{x=0} = a_i .$$

Suppose such a solution $y(x)$ exists, and let $y_{k_1 \ldots k_r}$ denote

$$\frac{\partial^r y}{\partial x^{k_1} \ldots \partial x^{k_r}}.$$

Then

$$x^{l_p} \frac{\partial}{\partial x^{l_p}} \ldots x^{l_1} \frac{\partial}{\partial x^{l_1}} (y_{jk} - y\, A_{jk}) x^j x^k = 0,$$

where p is any positive integer. From these identities we have

$$\left\{ y_{jkl_1 \ldots l_p} - \frac{\partial^p}{\partial x^{l_1} \ldots \partial x^{l_p}} (yA_{jk}) \right\} x^j \ldots x^{l_p} = 0,$$

and so

$$\left\{ y_{jkl_1 \ldots l_p} - [yA_{ijk}, l_1 \ldots l_p) \right\}_{x=0} = 0, \tag{2.2}$$

where the comma denotes partial differentiation, and $p!\, B_{(k_1 \ldots k_p)}$ stands for the sum of the quantities obtained by permuting the indices in all possible ways.

Let each set of numbers, except $a_1 \ldots a_n$ which are arbitrary,

$$a_0, a_j, \ldots, a_{jkl_1 \ldots l_p}, \ldots$$

be derived from those preceding it by the equations (2.2). If the series

$$y = a_0 + \sum_{p=1}^{\infty} \frac{1}{p!} a_{k_1 \ldots k_p} x^{k_1} \ldots x^{k_p} \tag{2.3}$$

converges for some value of x^i other than zero, it will represent a solution of the equations (2.1), which is regular near $x = 0$, and which will be uniquely determined by $a_0 \ldots a_n$. The lemma will, therefore, be established if only we can show that the series (2.3) converges.

Expanding A_{jk} as a power series about $x = 0$, we have

$$A_{jk} = A^0_{jk} + \sum_{p=1}^{\infty} \frac{1}{p!} A^0_{jkl_1 \ldots l_p} x^{l_1} \ldots x^{l_p}.$$

There exists, therefore, a number M such that

$$\left| \frac{1}{p!} A^0_{jkl_1 \ldots l_p} \right| < \frac{M}{\delta^p} \tag{2.4}$$

for all positive integral values of p. In the expansion of

$$\frac{\partial^r uv}{\partial x^{p_1} \ldots \partial x^{k_r}}$$

there will be $\dfrac{r!}{p!\,(r-p)!}$ terms in which u and v are differentiated, respectively, p and $(r-p)$ times, i.e., terms of the form

$$\frac{\partial^p u}{\partial x^{k_{a_1}} \ldots \partial x^{k_{a_p}}} \cdot \frac{\partial^{r-p} v}{\partial x^{k_{a_{p+1}}} \ldots \partial x^{k_{a_r}}},$$

where $a_1 \ldots a_r$ are the integers $1, \ldots r$. If we write λ_m for the greatest of the moduli $|a_{k_1 \ldots k_m}|$, where $a_{k_1 \ldots k_m}$ are defined by (2.2), we have, from (2.2) and (2.4)

$$\begin{aligned} \lambda^{m+2} < M\Big\{ &\frac{m!}{\delta^m}\lambda_0 + \frac{m!}{1!\,(m-1)!} \cdot \frac{(m-1)!}{\delta^{m-1}}\lambda_1 + \ldots \\ &+ \frac{m!}{p!(m-p)!} \cdot \frac{(m-p)!}{\delta^{m-p}}\lambda_p + \ldots + \lambda_m \Big\} \\ &= \frac{M\,m!}{\delta^m}\left(\lambda_0 + \lambda_1\delta + \frac{\lambda_2\delta^2}{2!} + \ldots + \frac{\lambda_m\delta^m}{m!}\right). \end{aligned}$$

Writing these inequalities

$$u_{m+2} < \frac{M\,\delta^2}{m+2} \cdot \frac{u_0 + \ldots + u_m}{m+1}$$

where $u_q = \dfrac{\lambda_q\,\delta^q}{q!}$, we have

$$u_{m+2} < \frac{M\,\delta^2}{m+2}\,\bar{u}_m$$

where $\bar{u}_m$ is the greatest of $u_0, \ldots, u_m$. If we chose an integer m_0 so that

$$\frac{M\,\delta^2}{m_0+2} < |, \text{ we have}$$

$$\begin{cases} u_{m_0+1} < \bar{u}_{m_0-1} \leqslant \bar{u}_{m_0} \\ u_{m_0+2} < \bar{u}_{m_0}. \end{cases}$$

Hence $\overline{u}_{m_0} = \overline{u}_{m_0+1} = \ldots = \overline{u}_{m_0+q} = \ldots,$
and so

$$u_N < K \tag{2.5}$$

for a suitable choice of K, and every positive integer N. From the definition of u_m and from (2.5), it follows that

$$\frac{\lambda_m}{m!} < \frac{K}{\delta_m},$$

or that the power series expansion of

$$y(x) = \frac{K}{1 - \dfrac{x^1 + \ldots + x^n}{\delta}}$$

dominates the series (2.3), which converges, therefore, for $|x^i| < \dfrac{\delta}{n}$. Hence the function $y(x)$, given by (2.3), satisfies the conditions of the lemma.

3. We shall now return to our problem. We are given a representation $x + x^0$ and are setting out to prove the existence of a representation, $z + z^0$, which is uniquely defined by the conditions

$$\begin{cases} P^\alpha_{jk}\, z^j z^k = 0 \\ \left(\dfrac{\partial z^\alpha}{\partial x^\beta}\right)_{x^\alpha = q^\alpha} = \delta^\alpha_\beta \\ z^\alpha = 0 \text{ for } x^\alpha = q^\alpha, \end{cases} \tag{3.1}$$

where q^i are the coördinates of a point, q^0 a value of the factor, and $P^\alpha_{\beta\gamma}$ are the components of the connection, Π, in $z + z^0$.

In case Π were flat, the differential equations

$$\frac{\partial^2 Z}{\partial x^\beta \partial x^\gamma} - \frac{\partial Z}{\partial x^\alpha}\,\Pi^\alpha_{\beta\gamma} = 0 \tag{3.2}$$

would be completely integrable, and would yield solutions, $Z^0, \ldots Z^n$, such that

$$\left(\frac{\partial Z^\alpha}{\partial x^\beta}\right)_q = \delta^\alpha_\beta \tag{3.3}$$

for $x^\alpha = q^\alpha$. From the condition $\Pi^\alpha_{\beta_0} = \delta^\alpha_\beta$ we have

$$Z^\alpha = e^{x_0 - q_0}\,\theta^\alpha(x),$$

were $\theta^\alpha(x)$ depend only on $x^1, \ldots x^n$. Let $P^\alpha_{\beta\gamma}$ be the components of π in the representation given by

$$\begin{cases} z^0 = \log Z^0 \\ \quad = x^0 - q^0 + \log \theta^0 \\ z^i = \dfrac{Z^i}{Z^0} \\ \quad = \dfrac{\theta^i}{\theta^0}. \end{cases} \tag{3.4}$$

Then we have

$$\begin{cases} Z^i = e^{z^0} z^i \\ Z^0 = e^{z^0}. \end{cases} \tag{3.5}$$

From (3.3) and (3.5) it follows that $Z^\alpha = \delta^\alpha_0$ and $\dfrac{\partial Z^\alpha}{\partial x^\beta} = \delta^\alpha_\beta$ for $x^\alpha = q^\alpha$. Since the expressions on the left-hand side of the equations (3.2) are the components of a projective tensor, we have

$$\frac{\partial^2 Z^\alpha}{\partial z^\beta \partial z^\gamma} - \frac{\partial Z^\alpha}{\partial Z^\sigma} P^\sigma_{\beta\gamma} = 0, \tag{3.6}$$

where Z^α are given in terms of $z + z^0$ by (3.5). It follows that

$$P^\alpha_{jk} = 0. \tag{3.7}$$

In the general case we shall establish the existence of solutions to (1.3) which satisfy (1.4), and make the same use of projective invariance to obtain the conditions (3.1). In the equations (1.3) we may change to any representation, provided we know its relation to $x + x^0$. We shall start with the representation $y + y^0$, where $y^0 = x^0 - q^0$, and y^i are affine normal coördinates for Π^i_{jk}, the point q, and the coördinate system x. For $x^\alpha = q^\alpha$ we have, therefore,

$$y^\alpha = 0, \quad \frac{\partial y^\alpha}{\partial x^\beta} = \delta^\alpha_\beta.$$

Let $\overline{\Pi}^\alpha_{\beta\gamma}$ be the components of π in $y + y^0$. Then in $y + y^0$ the components V^i are, by definition, y^i. We have, therefore.

$$\overline{\Pi}^i_{jk} V^j V^k = \overline{\Pi}^i_{jk} y^j y^k = 0,$$

and the equations (1.2) will become

$$\left(\frac{\partial^2 \theta}{\partial y^j \partial y^k} - \theta \overline{\Pi}_{jk} \right) y^j y^k = 0. \tag{3.8}$$

By our lemma we know that these equations admit a unique set of solutions $\theta^0, \ldots, \theta^n$, such that, writing $Z^\alpha = e^{y^0} \theta^\alpha$,

$$\left(\frac{\partial Z^\alpha}{\partial x^\beta}\right)_q = \left(\frac{\partial Z^\alpha}{\partial y^\beta}\right)_0 = \delta^\alpha_\beta. \tag{3.9}$$

We also have $Z^\alpha_{;0;\gamma} = Z^\alpha_{;\beta;0} = 0$, and therefore Z^α satisfy the equations

$$Z^\alpha_{;\beta;\gamma} V^\beta V^\gamma = 0.$$

Let $P^\alpha_{\beta\gamma}$ and $\overline{V}_\alpha$ be the components of Π and V in the representation given by

$$\begin{cases} Z^0 = \log Z^0 \\ \quad = x^0 - q^0 + \log \theta^0 \\ z^i = \dfrac{Z^i}{Z^0} \\ \quad = \dfrac{\theta^i}{\theta^0}. \end{cases} \tag{3.10}$$

Then by the same argument as that used in the flat case

$$P^\alpha_{jk} \overline{V}^j \overline{V}^k = 0. \tag{3.11}$$

To find the equations, in z, to the paths through q, we have to solve

$$\frac{dz^i}{d\sigma} = \overline{V}^i. \tag{3.12}$$

But on any path we have

$$\left(\frac{d^2 z^i}{d\sigma^2} + P^i_{kl} \frac{dz^k}{d\sigma} \frac{dz^l}{d\sigma}\right) \frac{dz^j}{d\sigma} = \left(\frac{d^2 z^j}{d\sigma^2} + P^j_{kl} \frac{dz^k}{d\sigma} \frac{dz^l}{d\sigma}\right) \frac{dz^i}{d\sigma},$$

and so from (3.11) and (3.12), we have, on the paths through q,

$$d^2 z^i dz^j = d^2 z^j dz^i$$

On these paths, therefore, $\dfrac{dz^i}{d\sigma} = z^i f(\sigma)$, and so $\overline{V}^i$ are proportional to z^i. The conditions (3.11) may, therefore, be written

$$P^\alpha_{jk} z^j z^k = 0,$$

and from (3.9) and (3.10), we have

$$\left(\frac{\partial z^\alpha}{\partial x^\beta}\right)_q = \delta^\alpha_\beta \text{ and } z^\alpha = 0, \text{ for } x^\alpha = q^\alpha.$$

The conditions (3.1), therefore, are satisfied by $z + z^\circ$.

The functions θ^α are the only solutions to (3.8) which satisfy (3.9), and from (3.10) it follows that the representation $z + z^0$ is uniquely determined by the conditions (3.1).

ON THE UNIFIED FIELD THEORY. I

BY TRACY YERKES THOMAS

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated September 30, 1930

In a number of notes in the Berlin *Sitzungsberichte*[1] followed by a revised account in the *Mathematische Annalen*,[2] Einstein has attempted to develop a unified theory of the gravitational and electromagnetic field by introducing into the scheme of Riemann geometry the possibility of distant parallelism. According to this view, there exists in each point of the underlying continuum of the world of space and time, a local cartesian coordinate system in which the Pythagorican theorem is satisfied. These local coördinate systems are determined by four independent vector fields with components $h_i^a(x)$ depending on the coördinates x^i of the continuum. There is, therefore, associated with each point, a configuration consisting of four independent vectors; it is assumed that these vector configurations are in parallel orientation in such a way that the arbitrary orientation of the configuration at one point determines uniquely the orientation of the configurations at all points of the continuum. This affords the possibility of determining whether or not two vectors at different points are parallel, namely, by comparing their components in the local systems: Two vectors are parallel if the corresponding components are equal when referred to a local system of coördinates. It is in the theory of the space of distant parallelism that Einstein has hoped to find his long-sought unification of electricity and gravitation.

It is important to determine an exact relation between the coördinates z^i of the local system and the coördinates x^i of the space of distance parallelism. This problem is solved in the present communication in such a way that certain requirements specified precisely in Sect. 2 are satisfied. As so defined there is a certain analogy between the local coördinates and the normal coördinates introduced into the theory of relativity by Birkhoff[3] and later discussed by the writer.[4] The relation between the z^i and x^i coördinates enables us to construct a set of *absolute invariants* with respect to transformations of the x^i coördinates sufficient for the complete characterization of the space of distant parallelism. Equations in the local system, in which the coördinates z^i are interpretable as coördinates of time and space, can be transformed directly into equations of general invariantive character. In view of this property we are led to the construction of a *system of wave equations* as the equations of the combined gravitational and electromagnetic field. This system is composed of 16 equations for the determination of the 16 quantities h_i^a and is closely analogous to the system of 10 equations for the determination of the 10 components

$g_{\alpha\beta}$ in the original theory of gravitation. It is an interesting fact that the covariant components h^i_α of the fundamental vectors, when considered as electromagnetic potential vectors, satisfy in the local coördinate system the universally recognized laws of Maxwell for the electromagnetic field in free space, as a consequence of the field equations.

It is my intention to supplement the results of the present paper by a series of papers devoted to an existence-theoretic treatment of the field equations and related problems.

1. Let us denote by $h^\alpha_i(x)$ for $i = 1,2,3,4$, the components of four contravariant vector fields; we observe that the Latin letter i is thus used to denote the vector, and the Greek letter α to denote the component of the vector. In general, we shall adopt this convention of Einstein, i.e., we shall employ a Latin letter for an index which is of invariantive character with respect to the arbitrary transformations of the x^i coördinates, and a Greek letter in all contrary cases. Departures from this rule, as well as departures from the rule that an index which appears twice in a term is to be summed over the values of its range, will be such as are easily recognized on observation. By means of the quantities $h^\alpha_i(x)$ we impose on the underlying continuum its structure as a space of distant parallelism in accordance with the following

POSTULATES OF SPACE STRUCTURE

A. *In any coördinate system* (x) *there exists a unique set of components* $\Delta^\alpha_{\beta\gamma}$ *of affine connection, for the determination of the affine properties of the continuum.*

B. *In any coördinate system* (x) *there exists a unique quadratic differential form* $g_{\alpha\beta}dx^\alpha dx^\beta$ *of signature* -2, *for the determination of the metric properties of the continuum.*

C. *At each point* P *of the continuum there is determined a configuration consisting of four orthogonal unit vectors issuing from* P.

D. *Corresponding vectors in the configurations, determined at two points* P *and* Q *of the continuum, are parallel.*

E. *The components* h^α_i *of the vectors determining the configuration at any point* P *of the continuum, are analytic functions of the* x^i *coördinates.*

The quadratic differential form in Postulate B will be referred to as the *fundamental form* and its coefficients $g_{\alpha\beta}$ as the components of the *fundamental metric tensor*. The four unit vectors with components $h^\alpha_i(x)$, which enter in Postulates C, D and E, will be referred to as the *fundamental vectors*. Since these vectors are orthogonal by Postulate C, the expression $g_{\alpha\beta}h^\alpha_i h^\beta_k$ is equal to zero for $i \neq k$; the condition that the fundamental vectors are unit vectors, likewise specified by Postulate C, means that for $i = k$, the expression $g_{\alpha\beta}h^\alpha_i h^\beta_k$ has the value ± 1. Hence we can write

$$g_{\alpha\beta}h_i^\alpha h_k^\beta = e_i\delta_k^i \tag{1.1}$$

where, following Eisenhart,[5] the convenient notation e_i for $+1$ or -1, is introduced. The fact that the fundamental form has signature -2 by Postulate B leads us to take

$$e_1 = 1, e_2 = e_3 = e_4 = -1;$$

more precisely, the signature -2 requires that one of the e_i have the value $+1$ and that the remaining e_i have values -1, the selection of the particular e_i to which the value $+1$ is assigned being an unessential matter of notation. From (1.1) we see that the product of the determinants $|g_{\alpha\beta}|$ and $|h_i^\alpha|^2$ is equal to -1. This shows that the determinant $|g_{\alpha\beta}|$ is *negative* and leads to the conclusion of the *independence* of the fundamental vectors. It is, therefore, possible to deduce from the fundamental vectors a system of four covariant vectors with components $h_\alpha^i(x)$ uniquely defined by the relations

$$h_i^\alpha h_\beta^i = \delta_\beta^\alpha, \quad h_k^\alpha h_\alpha^i = \delta_k^i;$$

the components h_i^α will be called the covariant components of the fundamental vectors to distinguish them from the contravariant components h_i^α of these vectors. In consequence of the above relations, equations (1.1) can be solved so as to obtain

$$\boxed{g_{\alpha\beta} = \sum_{i=1}^{4} e_i h_\alpha^i h_\beta^i} \tag{1.2}$$

as the equations defining the components of the fundamental metric tensor.

The condition that the fundamental vectors be parallel as demanded by Postulate D has its analytical expression in the equations

$$\frac{\partial h_i^\alpha}{\partial x^\gamma} + h_i^\beta \Delta_{\beta\gamma}^\alpha = 0.$$

This gives

$$\boxed{\Delta_{\beta\gamma}^\alpha = h_i^\alpha \frac{\partial h_\beta^i}{\partial x^\gamma}} \tag{1.3}$$

as the equations which define the components $\Delta_{\beta\gamma}^\alpha$ of affine connection.

Let us now suppose the existence of another system of fundamental vectors in the same space of distant parallelism. Let us denote the contravariant components of these vectors by ${}^*h_i^\alpha(x)$ in the (x) coördinate system, and let us put

$$*h_i^\alpha = a_i^k h_k^\alpha; \tag{1.4}$$

these equations define the quantities a_i^k as analytic functions of the x^i coördinates in consequence of Postulate E. Multiplying both members of (1.4) by $*h_\beta^i h_\alpha^j$ we obtain the equations

$$h_\beta^j = a_k^j *h_\beta^k \tag{1.5}$$

which represent the transformation induced by (1.4) on the covariant components h_α^i. On account of the uniqueness of determination of the components $\Delta_{\beta\gamma}^\alpha(x)$ demanded by Postulate A, we must have

$$h_i^\alpha \frac{\partial h_\beta^i}{\partial x^\gamma} = *h_i^\alpha \frac{\partial *h_\beta^i}{\partial x^\gamma}.$$

From (1.4) and these latter equations, it readily follows that *the quantities a_i^k are constants*. The quantities a_i^k are, however, not arbitrary constants, since they must satisfy a condition of orthogonality, namely,

$$\boxed{\sum_{i=1}^{4} e_i a_k^i a_l^i = e_k \delta_l^k} \tag{1.6}$$

which is obtained from (1.5) as the direct result of the uniqueness of determination of the fundamental form specified by Postulate B. Taking the determinant of both members of (1.6) we find that the determinant $|a_i^k|$ has the value ± 1. We can therefore define uniquely a set of quantities b_k^i by the equations

$$a_j^i b_k^j = \delta_k^i, \; a_k^j b_j^i = \delta_k^i.$$

Another form of the conditions (1.6) which is sometimes useful, can be derived in the following manner. Multiply both members of (1.6) by b_m^k so as to obtain

$$e_m a_l^m = e_l b_m^l,$$

or

$$a_l^m = e_l e_m b_m^l.$$

When we multiply both members of these latter equations through by $e_l a_l^k$ and sum of the index l, we find

$$\boxed{\sum_{l=1}^{4} e_l a_l^k a_l^m = e_k \delta_m^k} \; . \tag{1.7}$$

The transformation (1.4) in which the coefficients a_i^k are constants satisfying (1.6) or (1.7) will be called an *orthogonal transformation* of the components of the fundamental vectors.

2. The idea of the local coördinate system (z) is inherent in the idea of the configuration formed by the four vectors whose components h_i^α are subject to the orthogonal transformations (1.4), and, in fact, the construction of the local system was not carried beyond this stage by Einstein. Nevertheless a configuration of four vectors is not a coördinate system and we must face the problem of showing exactly how a system of local coordinates z^i can be defined. We require of the local system that it satisfy certain conditions which are stated precisely in the following

POSTULATES OF THE LOCAL SYSTEM

A. *With each point P of the space of distant parallelism there is associated a local coördinate system (z) having its origin at the point P.*

B. *The coördinate axes of the local system (z) at the point P are tangent to the fundamental vectors at P, in such a way that the z^i axis is tangent to the vector with components h_i^α and has its positive direction along the direction of this vector.*

C. *The interval ds is given by*

$$\boxed{ds^2 = (dz^1)^2 - (dz^2)^2 - (dz^3)^2 - (dz^4)^2} \tag{2.1}$$

at the origin of the local system.

D. *The paths[6] of the space of distant parallelism which pass through the origin of the local system (z) have the form*

$$\boxed{z^i = \xi^i v}, \tag{2.2}$$

where the ξ^i are constants and v is a parameter.

The above postulates give a complete geometrical characterization of the coördinates z^i of the local system. The *paths* which enter in Postulate D are, in general, defined as those curves which are given as solutions of the invariant system of equations

$$\frac{d^2x^\alpha}{dv^2} + \Lambda^\alpha_{\beta\gamma}\frac{dx^\beta}{dv}\frac{dx^\gamma}{dv} = 0, \tag{2.3}$$

where

$$\Lambda^\alpha_{\beta\gamma} = \frac{1}{2}(\Delta^\alpha_{\beta\gamma} + \Delta^\alpha_{\gamma\beta}); \tag{2.4}$$

such curves are generated by continuously displacing a vector parallel to itself along its own direction and are analogous to the straight lines of affine Euclidean geometry. If we denote the components of affine con-

nection $\Lambda^{\alpha}_{\beta\gamma}$ by $\lambda^{i}_{jk}(z)$, when referred to the local coördinate system, we have in consequence of Postulate D that

$$\lambda^{i}_{jk}\,\xi^{j}\,\xi^{k} = 0$$

along a path through the origin of the local system. It follows immediately from these latter equations that the equations

$$\lambda^{i}_{jk} z^{j} z^{k} = 0 \tag{2.5}$$

are satisfied identically in the local system. When we replace the components λ^{i}_{jk} in (2.5) by their values in terms of the components $\Lambda^{\alpha}_{\beta\gamma}$ we obtain a system of partial differential equations for the determination of the coördinates x^{i} in terms of the coördinates z^{i}, namely,

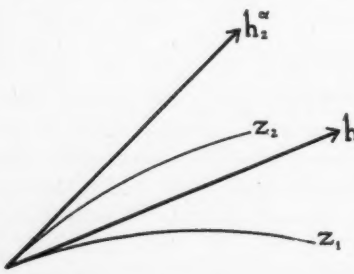
$$\left(\frac{\partial^2 x^{\alpha}}{\partial z^{j}\partial z^{k}} + \Lambda^{\alpha}_{\beta\gamma}\frac{\partial x^{\beta}}{\partial z^{j}}\frac{\partial x^{\gamma}}{\partial z^{k}}\right) z^{j} z^{k} = 0. \tag{2.6}$$

This system of equations possesses a unique solution $x^{\alpha} = \varphi^{\alpha}(z)$ satisfying a set of initial conditions

$$x^{\alpha} = p^{\alpha} \quad (z^{i} = 0), \tag{2.7}$$

$$\frac{\partial x^{\alpha}}{\partial z^{i}} = p^{\alpha}_{i} \quad (z^{i} = 0), \tag{2.8}$$

where p^{α} and p^{α}_{i} are arbitrary constants.[7] By Postulate A condition (2.7) is satisfied, provided that the constants p^{α} denote the coördinates of the point P; it remains to determine the values of the constants p^{α}_{i} in (2.8). For this purpose we consider the relation, imposed by Postulate B, between the local coördinates and the fundamental vectors at the point P; this is illustrated in the accompanying figure in which we have indicated only the positive z^{1} and z^{2} axes and the corresponding tangent vectors with components h^{α}_{1} and h^{α}_{2}, respectively. In consequence, we see that



$$\left(\frac{\partial x^{\alpha}}{\partial z^{i}}\right)_{z=0} = \sigma_{i} h^{\alpha}_{i}(p),$$

where the σ_{i} are positive constants. On account of this set of equations and the condition imposed by Postulate C, it readily follows that the σ_{i} are equal to ± 1. Hence $\sigma_{i} = +1$, since these constants are positive, and the above set of equations becomes

$$\left(\frac{\partial x^\alpha}{\partial z^i}\right)_{z=0} = h_i^\alpha(p). \tag{2.9}$$

The values $h_i^\alpha(p)$ are therefore to be ascribed to the above constants p_i^α. By repeated differentiation of (2.6) and use of the initial conditions (2.7) and (2.9), we determine the successive coefficients $H(p)$ of the power series expansions

$$x^\alpha = p^\alpha + h_i^\alpha(p)z^i - \frac{1}{2!}H_{ij}^\alpha(p)z^iz^j - \frac{1}{3!}H_{ijk}^\alpha(p)z^iz^jz^k - \ldots. \tag{2.10}$$

Thus we have that

$$H_{ij}^\alpha = \Lambda_{\beta\gamma}^\alpha h_i^\beta h_j^\gamma,$$

where

$$\Lambda_{\beta\gamma}^\alpha = \frac{1}{2}h_i^\alpha\left(\frac{\partial h_\beta^i}{\partial x^\gamma} + \frac{\partial h_\gamma^i}{\partial x^\beta}\right), \tag{2.11}$$

and

$$H_{ijk}^\alpha = \Lambda_{\beta\gamma\delta}^\alpha h_i^\beta h_j^\gamma h_k^\delta,$$

where

$$\Lambda_{\beta\gamma\delta}^\alpha = \frac{1}{3}\left[\left(\frac{\partial\Lambda_{\beta\gamma}^\alpha}{\partial x^\delta} + \frac{\partial\Lambda_{\gamma\delta}^\alpha}{\partial x^\beta} + \frac{\partial\Lambda_{\delta\beta}^\alpha}{\partial x^\gamma}\right) - 2(\Lambda_{\sigma\beta}^\alpha\Lambda_{\gamma\delta}^\sigma + \Lambda_{\sigma\gamma}^\alpha\Lambda_{\delta\beta}^\sigma + \Lambda_{\sigma\delta}^\alpha\Lambda_{\beta\gamma}^\sigma)\right], \tag{2.12}$$

etc. The jacobian of (2.10) does not vanish since it is equal to the determinant $|h_i^\alpha(p)|$ at the origin of the local system; hence (2.10) possesses a unique inverse. Either the transformation (2.10), or its inverse, gives a unique defination of the local coördinate system.

3. Let us denote by $\bar{z}^i$ the coördinates of the local system determined by the point P and the components of affine connection $\bar{\Lambda}_{\beta\gamma}^\alpha(\bar{x})$ which result from the components $\Lambda_{\beta\gamma}^\alpha(x)$ by an analytic transformation T of the x^α coördinates. The relation between the z^i and $\bar{z}^i$ coördinates must then be such that

$$z^i = 0\ (\bar{z}^i = 0), \quad \frac{\partial z^i}{\partial\bar{z}^k} = \delta_k^i\ (\bar{z}^i = 0). \tag{3.1}$$

Also this relation must be such as to satisfy the system of equations

$$\left(\frac{\partial^2 z^i}{\partial\bar{z}^j\partial\bar{z}^k} + \lambda_{p\sigma}^i\frac{\partial z^p}{\partial\bar{z}^j}\frac{\partial z^q}{\partial\bar{z}^k}\right)\bar{z}^j\bar{z}^k = 0. \tag{3.2}$$

Hence

$$\boxed{z^i = \bar{z}^i}. \tag{3.3}$$

This follows from the fact that (3.3) satisfies (3.2) and the initial conditions (3.1), and that (3.2) possesses a unique solution which satisfies the conditions (3.1). We can therefore say that the *local coördinates* z^i *remain unchanged when the underlying coördinates* x^i *undergo an arbitrary analytic transformation* T. In a similar manner we can show that *when the fundamental vectors undergo an orthogonal transformation* (1.4), *the local coördinates* z^i *associated with any point* P *likewise undergo an orthogonal transformation, i.e., a linear homogeneous transformation*

$$\boxed{z^i = a^i_k z^k_*}, \tag{3.4}$$

which leaves the form

$$\sum_{i=1}^{4} e_i z^i z^i$$

invariant. The behavior of the coördinates z^i of the local system (1) under arbitrary analytic transformations of the x^i coördinates, and (2) under orthogonal transformations of the fundamental vectors as described by the above italicized statements, is of great importance for the development of the theory of relativity.

4. If we transform the components of a tensor to a system of local coordinates z^i and evaluate at the origin of this system, we obtain a set of quantities which are of the nature of *absolute invariants* with respect to transformations of the x^i coördinates. To prove this formally we shall find it convenient to denote a set of tensor components $T^{\alpha\ldots\beta}_{\mu\ldots\nu}(x)$ with respect to the (x) system, by $t^{|i\ldots j|}_{|k\ldots m|}$ in the (z) coördinate system, i.e., we shall adopt the convention that in the (z) coördinate system, Latin letters when enclosed by $|\ \ |$, correspond to indices of covariant or contravariant character. If we put

$$T^{i\ldots j}_{k\ldots m} = (t^{|i\ldots j|}_{|k\ldots m|})_{z=0},$$

then

$$T^{i\ldots j}_{k\ldots m}(x) = \overline{T}^{i\ldots j}_{k\ldots m}(\overline{x}),$$

in consequence of (3.3). The explicit expressions for these invariants are obtained by evaluating at $z^i = 0$ both members of the set of equations

$$t^{|i\ldots j|}_{|k\ldots m|} = T^{\alpha\ldots\beta}_{\mu\ldots\nu}\frac{\partial x^\mu}{\partial z^k}\cdots\frac{\partial x^\nu}{\partial z^m}\frac{\partial z^i}{\partial x^\alpha}\cdots\frac{\partial z^j}{\partial x^\beta}. \tag{4.1}$$

This gives

$$T^{i\ldots j}_{k\ldots m} = T^{\alpha\ldots\beta}_{\mu\ldots\nu}h^\mu_k\ldots h^\nu_m h^i_\alpha\ldots h^j_\beta.$$

Similarly, if we differentiate the components t any number of times and evaluate at the origin of the (z) system, we obtain a set of quantities, namely,

$$T^{i\ldots j}_{k\ldots m,\,p\ldots q} = \left(\frac{\partial^r t^{|i\ldots j|}_{|k\ldots m|}}{\partial z^p \ldots \partial z^q}\right)_{z=0},$$

each of which is an absolute invariant with respect to analytic transformations of the x^i coördinates. On account of this invariant property, the name *absolute derivatives* will be used to refer to these quantities. To derive the explicit formulae for any absolute derivative $T^{i\ldots j}_{k\ldots m,\,p\ldots q}$ we have merely to differentiate (4.1) with respect to $z^p \ldots z^q$ and evaluate at the origin of the (z) system. For example, the first absolute derivative of the covariant vector with components T_α is given by the formula

$$T_{k,p} = \left(\frac{\partial T_\beta}{\partial x^\gamma} - T_\alpha \Lambda^\alpha_{\beta\gamma}\right) h^\beta_k \, h^\gamma_p.$$

We next consider the absolute derivatives of the fundamental vectors, using the covariant components h^i_α of these vectors as the basis of discussions. Let us denote these components by $A^i_{|j|}$ when referred to the (z) coördinate system. Then

$$A^i_{|j|} = h^i_\alpha \frac{\partial x^\alpha}{\partial z^j}. \tag{4.2}$$

In general, the absolute derivatives are defined by the equations

$$h^i_{j,k\ldots m} = \left(\frac{\partial^s A^i_{|j|}}{\partial z^k \ldots \partial z^m}\right)_{z=0}. \tag{4.3}$$

We note that

$$A^i_{|j|} = h^i_\alpha h^\alpha_j = \delta^i_j \quad (z^i = 0); \tag{4.4}$$

also that

$$\boxed{h^i_{j,k} = \frac{1}{2}\left[\frac{\partial h^i_\alpha}{\partial x^\beta} - \frac{\partial h^i_\beta}{\partial x^\alpha}\right] h^\alpha_j h^\beta_k} \tag{4.5}$$

and

$$\boxed{\begin{aligned} h^i_{j,kl} = \Bigg[&\frac{\partial^2 h^i_\alpha}{\partial x^\beta \partial x^\gamma} - \frac{\partial h^i_\alpha}{\partial x^\sigma}\Lambda^\sigma_{\beta\gamma} - \frac{\partial h^i_\sigma}{\partial x^\beta}\Lambda^\sigma_{\alpha\gamma} \\ &- \frac{\partial h^i_\sigma}{\partial x^\gamma}\Lambda^\sigma_{\alpha\beta} - h^i_\sigma \Lambda^\sigma_{\alpha\beta\gamma}\Bigg] h^\alpha_j h^\beta_k h^\gamma_l \end{aligned}}. \tag{4.6}$$

The special formulae (4.5) and (4.6) will be important in our later work.

By (4.3) the absolute derivative $h^i_{j,k\ldots m}$ is symmetric in the indices $k\ldots m$. Hence, these quantities satisfy the identities

$$\boxed{h^i_{j,k\ldots m} = h^i_{j,p\ldots q}}, \tag{4.7}$$

where $p\ldots q$ denotes any permutation of the indices $k\ldots m$. To derive other identities satisfied by the $h^i_{j,k\ldots m}$ we observe that the equations

$$\left(\frac{\partial A^i_{|j|}}{\partial z^k}\right) z^j z^k = 0 \tag{4.8}$$

which are equivalent to (2.5), are satisfied identically in the (z) coördinate system. By repeated differentiation of (4.8) and evaluation at the origin of the (z) system, we obtain

$$\boxed{S\,(h^i_{j,k\ldots m}) = 0}, \tag{4.9}$$

where the symbol S is used to stand for the summation of all the terms obtainable from the one in parenthesis by permutating the indices $jk\ldots m$ cyclically. As special cases of the identities (4.9) let us observe that

$$h^i_{j,k} + h^i_{k,j} = 0 \tag{4.10}$$

and

$$h^i_{j,kl} + h^i_{k,lj} + h^i_{l,jk} = 0. \tag{4.11}$$

The identities (4.7) *and* (4.9) *constitute a complete set of identities of the absolute derivatives* $h^i_{j,k\ldots m}$, in the sense that any other identity satisfied by these quantities is derivable from these identities.[8]

When we make an orthogonal transformation (1.4) of the components of the fundamental vectors, the components $A^i_{|j|}$ go over into a set of components ${}^*A^i_{|j|}$ referred to the (z_*) system, which are related to the $A^i_{|j|}$ by

$${}^*A^i_{|j|}\, a^p_i = A^p_{|q|} a^q_j; \tag{4.12}$$

this follows from (1.5), (3.4), and (4.2). Differentiating both members of (4.12) with respect to $z^k_*\ldots z^m_*$ and evaluating at the origin of the local system, we obtain

$${}^*h^i_{j,k\ldots m}\, a^p_i = h^p_{q,r\ldots t}\, a^q_j \ldots a^t_m. \tag{4.13}$$

We express this result by saying that the *absolute derivatives* $h^i_{j,k\ldots m}$ *constitute the components of a tensor with respect to orthogonal transformations of the fundamental vectors.* A similar discussion can, of course, be made on the basis of the contravariants components h^α_i of the fundamental vectors; also equations similar to (4.13) give the transformation of the components $T^{i\ldots j}_{k\ldots m}$ or more generally of the components $T^{i\ldots j}_{k\ldots m,p\ldots q}$, induced by the transformation (1.4) of the fundamental vectors.[9]

5. It is the sense of the local coördinate system (z) that the coördinate

z^1 is interpretable as the time coördinate and z^2, z^3, z^4, as rectangular cartesian coördinates of space. Let us, therefore, put

$$z^1 = t,\ z^2 = x,\ z^3 = y,\ z^4 = z,$$

in accordance with the unusual designations. More precisely, we should say that z^1 has the significance of a time coördinate and z^2, z^3, z^4, the significance of rectangular cartesian coördinates x, y, z *at the origin of coordinates*, i.e., in the infinitesimal neighborhood of the origin, since the interval ds has the exact form (2.1) only at the origin of the local system. In conformity with this interpretation of the coördinates of the local system, we now impose on the space of distant parallelism, the following

POSTULATE OF THE UNIFIED FIELD

The equations

$$\boxed{\frac{\partial^2 A^i_{\{j\}}}{\partial x^2} + \frac{\partial^2 A^i_{\{j\}}}{\partial y^2} + \frac{\partial^2 A^i_{\{j\}}}{\partial z^2} - \frac{\partial^2 A^i_{\{j\}}}{\partial t^2} = 0} \qquad (5.1)$$

are satisfied at the origin of the local system.

According to the above postulate, the field equations of the combined gravitational and electromagnetic field are of the nature of a system of wave equations—a type of equation which has already shown itself to be of fundamental importance for the study of gravitational or electromagnetic phenomena. The relationship of the above system of field equations to the field equations of the earlier theory of gravitation and to the Maxwell equations of the electromagnetic field is discussed in the following section.

When (5.1) is expressed in general invariative form, we have

$$\boxed{\sum_{k=1}^{4} e_k h^i_{j,kk} = 0} \qquad (5.2)$$

as the proposed system of field equations for the unified field theory. The system (5.2) is obviously invariant with respect to transformations of the x^i coördinates, since the factors which compose the system are themselves directly invariant; it is also invariant with respect to orthogonal transformations (1.4) of the fundamental vectors. To prove this we observe that

$$^*h^i_{j,kl} a^p_i = h^p_{q,rs} a^q_j a^r_k a^s_l$$

from (4.13). Putting l equal to k in these equations and multiplying through by e_k, we obtain

$$\left(\sum_{k=1}^{4} e_k {}^*h^i_{j,kk}\right) a^p_i = \left(\sum_{r=1}^{4} e_r h^p_{q,rr}\right) a^q$$

on making use of (1.7). This proves the assertion of invariance of (5.2) with respect to orthogonal transformations of the fundamental vectors.

6. If we denote the components of electric force by X, Y, Z and the components of magnetic force by α, β, γ then Maxwell's equations for the electromagnetic field in free space, are

$$\left\{\begin{aligned} \frac{\partial Z}{\partial y} - \frac{\partial Y}{\partial z} &= -\frac{\partial \alpha}{\partial t} \\ \frac{\partial X}{\partial z} - \frac{\partial Z}{\partial x} &= -\frac{\partial \beta}{\partial t} \\ \frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y} &= -\frac{\partial \gamma}{\partial t} \\ \frac{\partial \alpha}{\partial x} + \frac{\partial \beta}{\partial y} + \frac{\partial \gamma}{\partial z} &= 0 \end{aligned}\right. \tag{6.1}$$

and

$$\left\{\begin{aligned} \frac{\partial \gamma}{\partial y} - \frac{\partial \beta}{\partial z} &= \frac{\partial X}{\partial t} \\ \frac{\partial \alpha}{\partial z} - \frac{\partial \gamma}{\partial x} &= \frac{\partial Y}{\partial t} \\ \frac{\partial \beta}{\partial x} - \frac{\partial \alpha}{\partial y} &= \frac{\partial Z}{\partial t} \\ \frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z} &= 0, \end{aligned}\right. \tag{6.2}$$

where x, y, z are rectangular cartesian coördinates and t is the time. It is possible to write the above equations in a more contracted form. For this purpose we define a set of skew-symmetric quantities F_{jk} as the elements of the matrix

$$\left\| \begin{matrix} 0 & X & Y & Z \\ -X & 0 & -\gamma & \beta \\ -Y & \alpha & 0 & -\alpha \\ -Z & -\beta & \alpha & 0 \end{matrix} \right\|$$

and then construct the following two sets of equations

$$F_{jk,l} + F_{kl,j} + F_{lj,k} = 0 \tag{6.3}$$

and

$$\sum_{k=1}^{4} e_k F_{jk,k} = 0, \tag{6.4}$$

where $F_{jk,l}$ is now used to denote the ordinary partial derivative of F_{jk} with respect to z^l. The expanded forms of (6.3) and (6.4) give equations (6.1) and (6.2), respectively. Functions φ_j of the coördinates x, y, z and the time t are called electromagnetic potentials if they are such that

$$F_{jk} = \frac{\partial \varphi_j}{\partial z_k} - \frac{\partial \varphi_k}{\partial z^j}. \tag{6.5}$$

If the F_{jk} have the form (6.5), equations (6.3) are satisfied identically.

Now consider the covariant components h^i_α referred to the local system, i.e., the components $A^i_{|j|}$, and put

$$\varphi_j = A^{\ i}_{|j|} \tag{6.6}$$

for a fixed value of the index i; also define a set of quantities F_{jk} by (6.5), using the φ_j given by (6.6) for that purpose. Then

$$F_{jk} = \frac{\partial A^{\ i}_{|j|}}{\partial z^k} - \frac{\partial A^{\ i}_{|k|}}{\partial z^j}. \tag{6.7}$$

Differentiating and evaluating at the origin of the local system, we have

$$F_{jk,l} = h^i_{j,kl} - h^i_{k,jl};$$

the absolute derivatives $F_{jk,l}$ given by these equations satisfy the first set of Maxwell's equations (6.3) in consequence of the identities (4.11). Moreover,

$$\sum_{k=1}^{4} e_k F_{jk,k} = \sum_{k=1}^{4} (e_k h^i_{j,kk} - e_k h^i_{k,jk}) = \frac{3}{2} \sum_{k=1}^{4} e_k\, h^i_{j,kk}.$$

Hence, the second set of Maxwell's equations (6.4) is satisfied in consequence of the field equations (5.2). In other words, *the covariant components of the fundamental vectors when considered as electromagnetic potential vectors, satisfy, in the local coördinate system, the universally recognized laws of Maxwell for the electromagnetic field in free space, as a consequence of the field equations* (5.2). This fact strongly suggests that the components h^i_α will play the rôle of electromagnetic potentials in the present theory.

Taking account of the field equations (5.2) we easily deduce the system of equations

$$\sum_{k=1}^{4} e_k g_{ij,kk} = 2 \sum_{k,l=1}^{4} e_k\, e_l\, h^l_{i,k}\, h^l_{j,k}. \tag{6.8}$$

Or, denoting the components $g_{\alpha\beta}$ by $B_{|ij|}$ when referred to the local coordinate system, we have

$$\frac{\partial^2 B_{|ij|}}{\partial x^2} + \frac{\partial^2 B_{|ij|}}{\partial y^2} + \frac{\partial^2 B_{|ij|}}{\partial z^2} - \frac{\partial^2 B_{|ij|}}{\partial t^2} = -2 \sum_{k,l=1}^{4} e_k e_l h^l_{i,k} h^l_{j,k} \tag{6.9}$$

at the origin of the local system. Now in the earlier theory of gravitation, the field equations could be given a form analogous to (6.9) in which the right members were equal to zero;[3] since the nature of the solution of (6.8) is determined by the form of the left members of (6.9) it follows that the functions $g_{\alpha\beta}$ possess the same general character as in the earlier theory of gravitation. Moreover, the interpretation of the quantities h^i_α as electromagnetic potentials would lead us to expect that in a purely gravitational field, i.e., more precisely, in a comparatively inappreciable electromagnetic field, the square of the $h^i_{j,k}$ would be negligibly small quantities. In this case the right members of (6.9) vanish approximately and we are left with the system

$$\frac{\partial^2 B_{|ij|}}{\partial x^2} + \frac{\partial^2 B_{|ij|}}{\partial y^2} + \frac{\partial^2 B_{|ij|}}{\partial z^2} - \frac{\partial^2 B_{|ij|}}{\partial t^2} = 0$$

as a first approximation. It is, therefore, to be expected that the quantities $g_{\alpha\beta}$ will successfully assume the rôle of *gravitational potentials* as in the previous theory of gravitation.

[1] A. Einstein, *Berliner Berichte*, 1928–30.

[2] A. Einstein, "Auf die Riemann-Metrik und den Fern-Parallelismus gegründete einheitliche Feldtheorie," *Math. Ann.*, **102** (1930), pp. 685–697.

[3] G. D. Birkhoff, *Relativity and Modern Physics*, Harvard University Press (1923), pp. 124 and 228.

[4] T. Y. Thomas, "The Principle of Equivalence in the Theory of Relativity," *Phil. Mag.*, **48** (1924), pp. 1056–1068.

[5] L. P. Eisenhart, *Riemannian Geometry*, Princeton University Press (1926), Chap. II.

[6] This has reference to the *path* in the sense in which that word is used in "The Geometry of Paths," *Trans. Am. Math. Soc.*, **25** (1923), pp. 551–608.

[7] It is obvious that the system of differential equations (2.6) possesses a unique formal solution

$$x^\alpha = p^\alpha + p^\alpha_i z^i + \frac{1}{2!} p^\alpha_{ij} z^i z^j + \ldots \qquad (a)$$

such that the conditions (2.7) and (2.8) are satisfied. A proof of convergence of the series (*a*) can be given in the following manner. Consider a system of equations

$$\frac{\partial^2 X^\alpha}{\partial z^j \partial z^k} - F^\alpha_{\beta\gamma} \frac{\partial X^\beta}{\partial z^j} \frac{\partial X^\gamma}{\partial z^k} = 0, \qquad (b)$$

where the $F^\alpha_{\beta\gamma}$ are analytic functions of the variables X^α. The conditions of integrability of (*b*) are that

$$\frac{\partial F^\alpha_{\beta\gamma}}{\partial X^\delta} - \frac{\partial F^\alpha_{\beta\delta}}{\partial X^\gamma} + F^\alpha_{\mu\delta} F^\mu_{\beta\gamma} - F^\alpha_{\mu\gamma} F^\mu_{\beta\delta} = 0$$

identically. These conditions are satisfied by taking all the functions $F^\alpha_{\beta\gamma}$ equal to one another in accordance with the equations

$$F^{\alpha}_{\beta\gamma} = \frac{M}{1 - \dfrac{X^1 + \ldots + X^4}{\rho}},$$

where M and ρ are constants; the system (b) then possesses a unique solution

$$X^{\alpha} = P^{\alpha} + P^{\alpha}_{i}\, z^{i} + \frac{1}{2!} P^{\alpha}_{ij}\, z^{i}\, z^{j} + \ldots, \qquad (c)$$

satisfying the conditions

$$X^{\alpha} = P^{\alpha} \quad (z^{i} = 0),$$

$$\frac{\partial X^{\alpha}}{\partial z^{i}} = P^{\alpha}_{i} \quad (z^{i} = 0).$$

Now choose the above constants M and ρ so that each function $F^{\alpha}_{\beta\gamma}$ dominates the corresponding functions $\Lambda^{\alpha}_{\beta\gamma}$, and, furthermore, choose the P^{α} and P^{α}_{i} such that

$$P^{\alpha} \geqq | p^{\alpha} |, \quad P^{\alpha}_{i} \geqq | p^{\alpha}_{i} |.$$

In view of the fact that (c) is likewise a solution of the system

$$\left(\frac{\partial^2 X^{\alpha}}{\partial z^{j} \partial z^{k}} - F^{\alpha}_{\beta\gamma} \frac{\partial X^{\beta}}{\partial z^{j}} \frac{\partial X^{\gamma}}{\partial z^{k}} \right) z^{j} z^{k} = 0,$$

it is then easily seen that each expansion (c) dominates the corresponding expansion (a). The convergence of the expansions (a) within a sufficiently small neighborhood of the values $z^{j} = 0$, is therefore established.

[9] The proof is analogous to that given in my paper "The Identities of Affinely Connected Manifo'ds," *Math. Zeit.*, **25** (1926), pp. 714–722.

[9] A method of covariant differentiation based on the non-symmetric components of affine connection $\Delta^{\alpha}_{\beta\gamma}$ is used by Einstein.[3] The point of view adopted in the present investigation in which covariant differentiation, or more precisely absolute differentiation, is brought into relationship with the local coördinate system, necessarily depends on a symmetric connection; for this reason I have rejected the method of Einstein. It would be possible also to define covariant or absolute derivatives on the basis of the Christoffel symbols $\Gamma^{\alpha}_{\beta\gamma}$ derived from the components $g_{\alpha\beta}$ and, in fact, this has been advocated by T. Levi-Civita, "Vereinfachte Herstellung der Einsteinschen einheitlichen Feldgleichungen," *Berliner Berichte*, **1929**. This method for the construction of invariants can be brought into relationship with the local coördinates; in fact, it is only necessary to replace Postulate D of the *Postulates of the Local System* by a similar postulate on the *geodesics* of the space of distant parallelism. My primary objection to this is that we lose a good deal of the simplification inherent in the above theory. For example, in place of the quantities $h^{i}_{j,k}$ given by (4.5) we would have the more complicated set of invariants

$$\left(\frac{\partial h^{i}_{\alpha}}{\partial x^{\beta}} - \Gamma^{\sigma}_{\alpha\beta} h^{i}_{\sigma} \right) h^{\alpha}_{j} h^{\beta}_{k}$$

which do not possess the simple property of shew-symmetry enjoyed by the $h^{i}_{j,k}$, and which are therefore not so directly related to the electromagnetic forces (see Sect. 6).

We have then, in all, three different methods for the construction of invariants, namely, the method of Einstein, that of Levi-Civita, and the method used in the above

investigation. Now there is a certain psychological influence exerted by the method itself upon the investigator. What I mean is that one is naturally led to the construction of special invariants which, on the basis adopted for the formation of invariants, are represented by simple analytical expressions. So, for example, the field equations proposed by Einstein[2] have a very simple analytical form in terms of the covariant derivatives used by him, but these same equations would be considerably more complicated in form if expressed in terms of the covariant derivatives used by Levi-Civita; also the simple form (5.1) or (5.2) of the field equations assumed in the above investigation is peculiar to the method of absolute differentiation which I have adopted. Thus the different methods of construction of invariants will lead us, in practice, to the assumption of different systems of field equations; these systems of field equations must, roughly speaking, be of the same general character but will nevertheless, not be precisely equivalent, considered as systems of partial differential equations.

It is possible to develop a process by which the above methods are brought into relationship with one another and which will, moreover, permit the ready construction of invariants depending on combinations of these methods. This process has its geometrical foundation in the study of those surfaces $x^\alpha = f^\alpha(u, v)$ which are defined as solutions of the system of equations

$$\left\{\begin{aligned} &\frac{\partial^2 x^\alpha}{\partial u \partial v} + \Delta^\alpha_{\beta\gamma}\frac{\partial x^\beta}{\partial u}\frac{\partial x^\gamma}{\partial v} = 0, \\ &\frac{\partial^2 x^\alpha}{\partial u^2} + \Gamma^\alpha_{\beta\gamma}\frac{\partial x^\beta}{\partial u}\frac{\partial x^\gamma}{\partial u} = 0 \quad (v = 0), \\ &\frac{\partial^2 x^\alpha}{\partial v^2} + \Delta^\alpha_{\beta\gamma}\frac{\partial x^\beta}{\partial v}\frac{\partial x^\gamma}{\partial v} = 0 \quad (u = 0), \end{aligned}\right. \tag{a}$$

and so constitutes a generalization of the process of covariant differentiation or extension, as developed in *The Geometry of Paths*. In this way we are led to a set of relations $x^\alpha = g^\alpha(y, z)$ which, for z^i = const., denote a transformation to a system of coördinates y^i, and which for y^i = const. denote a transformation to a system of coördinates z^i. If $t(y, z)$ represents the components of a tensor either in the (y) or (z) coördinate system, then

$$\left(\frac{\partial^r t(y, z)}{\partial y^\alpha \ldots \partial y^\beta \partial z^\gamma \ldots \partial z^\delta}\right) y = 0, z = 0$$

defines, when considered as a function of the x^i coördinates, the components of a tensor in the (x) coördinate system. As an alternative method of procedure the system (a) can be replaced by the system

$$\left\{\begin{aligned} &\left(\frac{\partial^2 x^\alpha}{\partial y^j \partial z^k} + \Delta^\alpha_{\beta\gamma}\frac{\partial x^\beta}{\partial y^j}\frac{\partial x^\gamma}{\partial z^k}\right) y^j z^k = 0, \\ &\left(\frac{\partial^2 x^\alpha}{\partial y^j \partial y^k} + \Gamma^\alpha_{\beta\gamma}\frac{\partial x^\beta}{\partial y^j}\frac{\partial x^\gamma}{\partial y^k}\right) y^j y^k = 0 \quad (z = 0), \\ &\left(\frac{\partial^2 x^\alpha}{\partial z^j \partial z^k} + \Delta^\alpha_{\beta\gamma}\frac{\partial x^\beta}{\partial z^j}\frac{\partial x^\gamma}{\partial z^k}\right) z^j z^k = 0 \quad (y = 0). \end{aligned}\right. \tag{b}$$

The consideration of system (b) enables us, moreover, by imposing initial conditions corresponding to (2.9), to develop a theory of absolute invariants which is a generalization of that of the above investigation. The details of this process will not be developed here as it is not necessary for our work on the theory of relativity.

THE CRITICAL POINTS OF A FUNCTION OF n VARIABLES

BY MARSTON MORSE

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY

Communicated October 6, 1930

This paper contains among other results a treatment of the critical points of a real analytic function without restriction as to the nature of the critical points. Together with the author's unpublished results on the removal of the boundary conditions it constitutes a complete treatment of the problem, the first of its kind.

In the most of the paper the function considered is of class C'', and may have critical loci not even complexes, in fact, an infinite set of such loci. Moreover, even in the analytic case it is not assumed that the critical loci are complexes, a considerable advantage.

Let (x) be a point in Euclidian n-space in a finite region Σ, bounded by a closed set M, consisting of a finite number of connected, regular, non-intersecting $(n-1)$ spreads of class C''. Let $f(x)$ be a real function of (x) of class C'' on Σ, with a positive, exterior normal directional derivative on M. We assume the critical values of f are finite in number.

By a *critical set* g will be understood any closed set of critical points of f which yields one critical value, which includes all critical points to which it may be connected among critical points, and which is at a positive distance from other critical points. It may or may not be connected. In general,, it will not be a complex. Suppose we have a critical set at which $f = 0$.

Let N and N^1 be any two open neighborhoods of g of which N^1 is interior to N. By $\overline{N}$ and $\overline{N}^1$ we shall mean those points of N and N^1 at which $f<0$.

In the analytic case, and much more generally the following sets of cycles exist for an N^1 sufficiently near g.

(α). *A complete set* $(a)_k$ *of k-cycles on* $\overline{N}^1$, *independent on* $\overline{N}$, *dependent on* N^1.

(β). *A complete set* $(c)_k$ *of k-cycles on* N^1 *independent on* N *of the k-cycles on* $\overline{N}$. (Called critical cycles.)

Moreover, for neighborhoods N sufficiently near g and correspondingly small N^1, the number of cycles in these sets is independent of N and N^1. This we prove for cases much more general than the analytic.

We define the kth *type number* M_k of a critical set g to be the number of cycles in sets $(a)_{k-1}$ and $(c)_1$.

From its definition it appears that the type numbers are *invariant* under homeomorphisms that carry critical sets into critical sets. The type number of a finite ensemble of critical sets is the sum of the type numbers of the component sets.

We replace $(a)_{k-1}$ by an algebraically equivalent set composed of two

sets, a set $(l)_{k-1}$ of $(k-1)$-cycles that bound on $f < 0$, called *linkable*, and a set $(b)_{k-1}$ called *newly bounding*, no linear combination of whose members, not null, so bounds. Each member of $(l)_{k-1}$ bounds a k-chain on $f < 0$, and one on N. The difference of these two k-chains gives a k-cycle *linking* the member of $(l)_{k-1}$.

Let a and b be two ordinary values of f between which lies just one critical value zero, with a critical set g, not necessarily finite, or connected, or a complex. Let the domains $f \leqq a$ and $f \leqq b$ be denoted, respectively, by A and B. Let us call a k-cycle, on A, which is non-bounding on B, *invariant*. We have the following theorem whose proof, largely by means of deformations, takes less than two pages.

THEOREM. *A complete set of k-cycles for B is composed of complete sets of invariant, linking, and critical k-cycles, respectively (Mod. m Prime).*

We associate with g, m_k^+ and m_k^- ideal critical points of *increasing* and *decreasing* types, respectively, of which m_k^+ shall equal the number of linking and critical k-cycles in complete sets, and of which m_k^- shall equal the number of newly bounding $(k-1)$-cycles in a complete set.

Let a and b be any two non-critical constants. Between the differences ΔR_k, namely, the kth Betti number of B minus that of A, and the numbers M_k, m_k^+, m_k^-, summed over all critical sets on $B-A$ the earlier relations now hold, namely,

$$\Delta R_k = m_k^+ - m_{k+1}^- \qquad (k = 0, 1, \ldots, n)$$
$$M_k = m_k^+ + m_k^-$$

with $m_0^- = m_{n+1}^- = 0$. Important equalities are obtained by eliminating the numbers m_k^+. Similarly for m_k^-. A host of equalities and inequalities results.

We also prove that M_k will exceed or equal the number of k-cycles on B, independent of k-cycles on A, in a complete set, plus the number of $(k-1)$-cycles on A independent on A, but bounding on B, in a complete set.

We determine the type numbers more simply in terms of *neighborhood functions* $\varphi(x)$ related to each critical set g.

Such a function φ shall be of class C'' neighboring g, take on a proper relative minimum on g, and possess a gradient never parallel to that of f at points not on g at which $f = 0$.

We show that $\Sigma f_{x_i} f_{x_i}$ is always a neighborhood function in the analytic case. For maximizing or minimizing critical sets, $-f$ and f serve respectively as neighborhood functions *in any case whatsoever*. We prove the existence of functions φ in other cases.

Let φ_e and φ_e^0, respectively, denote the domains $\varphi = e$ and $0 < \varphi \leqq e$, where e is a positive constant less than a certain positive constant r. The points of these domains at which $f<0$ will be respectively denoted by $\bar{\varphi}_e$ and $\bar{\varphi}_e^0$. The domain $\varphi \leqq e$ will be denoted by φ_e^*.

We show that the domains $\bar{\varphi}_e$ are bounded by non-singular, closed, regular $(n-2)$-spreads, and are homeomorphic for each e. Each domain $\bar{\varphi}_e^0$ is deformable on itself without increasing φ onto any domain $\bar{\varphi}_\eta^0$ with $\eta < e$, keeping $\bar{\varphi}_\eta^0$ fixed. The domains φ_e^* are homeomorphic for each e.

The type number M_k is shown to be the number of cycles in the following complete sets.

(*a*) *A complete set of* $(k-1)$*-cycles on* $\bar{\varphi}_e$ *independent on* $\bar{\varphi}_e$*, bounding on* φ_e^*.

(*b*) *A complete set of k-cycles on* φ_e^* *independent on* φ_e^* *of the cycles on* $\bar{\varphi}_e$ (critical k-cycles).

We see that M_0 and M_n are null except when the critical set is respectively minimizing or maximizing, and in the later cases one for connected sets.

For a minimizing set on which $f = 0$ we can take $\phi = f$. We see that (*a*) is empty, and M_k is the kth Betti number of $f \leqq e$ neighboring g. For a maximizing set we take $\varphi = -f$. We see that (*b*) is empty in Euclidian n-space and $M_k = R_{n-k}$, where R_{n-k} is a Betti number of $f \geqq -e$. In the case of an isolated critical point, f analytic, we can take $\varphi = x_i x_i$ and Brown's definition is given by (*a*) and (*b*), so that his results on critical points all follow. The author's earlier results are, with more difficulty, included in this paper. Brown's results* in the plane also follow.

We justify our results further by showing that we can deform the function slightly, preserving boundary conditions and class, obtaining a function with non-degenerate critical points which are grouped in the respective neighborhoods of the critical sets of f. Moreover, in the neighborhood of *each connected critical set* of f the new critical points there found have type number sums at least as great as the corresponding type number of g.

In the analytic case the deformation is made analytically.

The author has just finished a proof that the type numbers of a general critical set always exist, and are finite. The critical values are assumed isolated only for points neighboring the given critical set.

* Brown, *Ann. Math.*, **31** (1930), p. 449. Here also are references to Birkhoff, Whyburn, Poincaré, Brown, Morse, and for analysis situs Alexander. See also

Morse, *Proc. Nat. Acad. Sci.*, **13** (1927), p. 813.

Morse, *Trans. Amer. Math. Soc.*, **36** (1930). In this paper is found the origin of the methods of the present paper.

The Colloquium Lectures by Lefchetz, not yet out at the time of writing, will undoubtedly be of value in this connection.